Development of a School Readiness Index for Canadian Preschoolers:

A Methodological Report and Time Series Analysis

Cameron N. McIntosh

Health Information and Research Division, Statistics Canada

and

Dafna E. Kohen

Health Information and Research Division, Statistics Canada

Department of Epidemiology and Community Medicine,

University of Ottawa

CANADIAN COUNCIL ON LEARNING CCL CCA CONSEIL CANADIEN SUR L'APPRENTISSAGE

Summary

Determining children's readiness for school is of considerable importance from both an educational research and policy perspective. The purpose of the current paper is twofold. In the first part, we describe the methodological steps used in the development of a statistical function to estimate school readiness scores for Canadian preschoolers; in the second part, we present our estimates in a provincial-level time series analysis for the years 1998, 2000, 2002, and 2004.

An initial conceptual model outlining five proposed domains of school readiness was used as a guide in assembling the data sources and carrying out multivariate statistical modeling: (i) Health and Physical Well-Being and Motor Development, (ii) Social and Emotional Development, (iii) Approaches to Learning, (iv) Language and Communication Skills, and (v) Cognition and General Knowledge. Information on these domains was obtained by pooling data on key school readiness variables administered to 4- and 5-year-olds in cycles 3 and 4 of the Canadian National Longitudinal Survey of Children and Youth (NLSCY). Longitudinal Follow-up data on school outcomes were obtained from cycle 6. Multiple imputation was used to replace large quantities of missing data prior to multivariate analysis, but technical problems were encountered in subsequent attempts to factor-analyze the imputed data sets, seemingly due to instability in the replaced values. Reducing the sample to only 4- and 5-year olds interviewed at cycle 3 and followed up at cycle 6, and using pairwise deletion rather than multiple imputation to accommodate missing data, allowed for the extraction of three underlying

factors from the collection of observed variables on readiness: Health, Social/Emotional Development, and Verbal Ability.

These three readiness factors were then considered as predictors of a School Outcomes factor in a structural equation model (SEM). In the final model, Verbal Ability and Social/Emotional Development were direct predictors of School Outcomes; whereas Health was a background factor directly influencing both Verbal Ability and Social/Emotional Development thus having an indirect effect on School Outcomes. This model provided a weighted linear additive function to generate a school readiness score for any respondent *i* from factor scores on Verbal Ability and Behaviour; formally:

$$\textbf{School Readiness}_i = 0.29*\textbf{Verbal Ability}_i + 0.28*\textbf{Social/Emotional Development}_i$$

Using the final model developed by the SEM, factor scores were estimated for Verbal Ability and Social/Emotional Development (adjusted for the effects of Health) for 4- and 5-year-olds, across NLSCY cycles 3 to 6. Applying the above prediction equation to each case yielded school readiness scores, which were converted into normal curve equivalents (NCEs) and presented in the form of a provincial-level time series analysis: 1998, 2000, 2002, and 2004. The NCEs allow for a relative evaluation of levels of school readiness; in the present case, one can examine the distribution of the provincial means to a fixed national mean. Interesting differences were found both within cycles and over time.

For example, in 1998, Newfoundland, Nova Scotia, and New Brunswick, and British Columbia were all significantly lower than the national mean. Newfoundland moved to being significantly greater than the national mean for all of the remaining years in the series. Nova Scotia shifted to the level of the national mean in 2000 through 2004, as did British Columbia. New Brunswick was equivalent to the national mean in 2000, greater in 2002, and then returned to the level of the national mean in 2004. Prince Edward Island was not significantly different from the national mean in any of the years examined. Ontario showed the highest degree of fluctuation, being significantly higher than the national mean in 1998, lower in 2000, higher in 2002, and then falling to the level of the national mean in 2004. Manitoba was significantly lower than the national mean in both 1998 and 2000, but not significantly from it in both 2002 and 2004. Saskatchewan was not significantly different than the national mean in 1998, was lower in both 2000 and 2002, but was at the level of the national mean again in 2004. Alberta was significantly higher than the national mean in 1998, not reliably different from the national mean in 2000, lower in 2002, and again not significantly different in 2004 Quebec was not significantly different from the national mean in 1998, but was significantly lower for the remainder of the time series. Examining the underlying reasons for these differences (e.g., differences in policy and programs) is a potentially promising avenue for future research.

Significant limitations of the current analysis include (i) the ultimately narrow coverage of the five school readiness domains identified in the original conceptual framework for the study, (ii) the lack of an absolute benchmark or criterion value against which to compare observed degrees of school readiness at the individual or aggreagate (i.e., provincial) level, and (iii) inconsistencies in measurement of the observed variables reflecting readiness domains across NLSCY cycles. Despite these shortcomings, we feel that the statistical methodology used in the current study is sound and provides a solid template for constructing national indexes from survey data. A crucial next step in addressing the information gaps in the present school readiness index is the exploration of other longitudinal, nationally representative data sets containing a broader array of variables on school readiness and school outcomes for a given cohort. Analysis of such data sets with the statistical techniques used here has the potential to greatly improve the breadth and depth of the information encapsulated in the school readiness index, and thereby produce a more reliable means of monitoring school readiness at the aggregate population level, for the purposes of both research and policy-making.

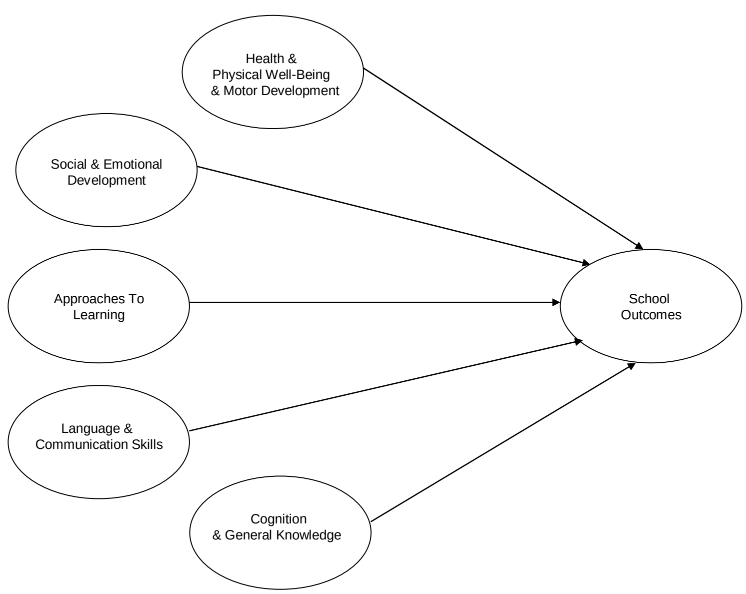
Development of a School Readiness Index for Canadian Preschoolers:

A Methodological Report and Time Series Analysis

Introduction

The current paper focuses on the creation of a school readiness indicator for Canadian preschoolers. The conceptual foundation for the indicator, articulated in previous articles (see Britto et al., 2006; Kohen, 2006, Kohen & Britto, 2006 for review), is summarized in Figure 1. As shown in Figure 1, the following five domains of school readiness are hypothesized to act in concert to influence children's school outcomes: (i) Health and Physical Well-Being and Motor Development, (ii) Social and Emotional Development, (iii) Approaches to Learning, (iv) Language and Communication Skills, and (v) Cognition and General Knowledge. School outcomes include academic performance, as well as various related attitudes and behaviours (e.g., enjoyment of school, discipline problems).

Figure 1: Guiding Conceptual Model Depicting Five School Readiness Domains As Simultaneous Predictors of School Outcomes. (Measured variables reflecting the domains are not shown here but rather described in the text).



Under the direction of this general conceptual framework, the present study uses data from the Canadian National Longitudinal Survey of Children and Youth (NLSCY; Statistics Canada/Human Resources Development Canada) to construct a composite index reflecting preschoolers' level of school readiness. For statistical modeling purposes, the diagram in Figure 1 can also be represented in the form of a regression equation, in which the overall school outcomes score for student *i* as a function of the five readiness domains is formally expressed as:

$$\textit{School Outcomes}_i = \beta_0 + \beta_1 * \textit{Health and Physical Well-Being and Motor Development}_i + \beta_2 * \textit{Social and Emotional Development}_i + \beta_3 * \textit{Approaches to Learning}_i + \beta_4 * \textit{Language and Communication Skills}_i + \beta_5 * \textit{Cognition and General Knowledge}_i + \varepsilon_i \qquad \textbf{[1]},$$

where β_0**, is the intercept term,** β_1 through β_5 are the regression coefficients representing the contribution of each readiness domain to school outcomes, and ε_i **represents unexplained variation. Once the coefficients are** estimated, a predicted value on school readiness for the *i*th case can be computed as a weighted linear composite of scores on the five domains, as follows:

$$\textit{School Readiness}_i = \hat{\beta}_1 * \textit{Health and Physical Well-Being and Motor Development}_i + \hat{\beta}_2 * \textit{Social and Emotional Development}_i + \hat{\beta}_3 * \textit{Approaches to Learning}_i + \hat{\beta}_4 * \textit{Language and Communication Skills}_i + \hat{\beta}_5 * \textit{Cognition and General Knowledge}_i \qquad \textbf{[2]},$$

where scores on each of the school readiness domains are in standardized form.

This paper is organized in two parts. The first describes the methodological steps used to estimate the parameters in equation 1 above, in order to derive a statistical function for computing individual-level school readiness scores from Canadian national data on the readiness domains (equation 2). Using multiple waves of NLSCY data, the second part of this paper presents a provincial-level time series analysis of school readiness – for the years 1998, 2000, 2002, and 2004 – for both descriptive and comparative purposes.

Part I: Methodology for Constructing a School Readiness Index

Overview

Construction of the proposed school readiness indicator entails a number of methodological steps, which are summarized here and then presented in more detail below. In the first step, cycles 3 (1998 - 1999), 4 (2000 - 2001), and 6 (2004 - 2005) of the NLSCY will be used to create a pooled sample of Canadian children aged 4 and 5, on variables reflecting both school readiness domains (cycles 3 and 4) and school outcomes (cycle 6). Combining observations from these three cycles on the study variables is necessary to address a number of data availability issues. The primary challenge is that while the cycle 3 respondents have longitudinal follow-up data for the school outcomes measured at cycle 6, cycle 3 alone provides only sparse coverage of the various domains of readiness identified above. In Cycle 4, a broader array of school readiness indicators was administered; however, these children have no school outcomes data for cycle 6, which is necessary to quantify the predictive ability of the school readiness domains and thereby derive a statistical formula for computing a summary index. Therefore, the second step will be to impute all missing values in the pooled data set, prior to any analyses.

The third step will consist of exploratory factor analysis (EFA) of the variables comprising the school readiness domains, using the imputed data sets. This strategy serves to reduce the broad array of observed domain indicators to a more parsimonious set of latent variables for predicting school outcomes. Although the theoretical background for the different readiness variables provides some rationale for assigning them to particular constructs, it seems most prudent to consider the entire set of readiness indicators as an initially theoretically-driven collection of variables, yet determine their optimal sub-groupings empirically. This step also deals with potential redundancy among indicators from different readiness domains.

However, the variables reflecting school outcomes will not be subjected to an EFA but rather all specified *a priori* as indicators of a single latent construct, in order to facilitate the construction of the school readiness index. Specifically, the index will be statistically defined as a weighted composite of the readiness domains, based on their independent contributions to a single school outcomes construct. Therefore, to estimate the optimal prediction weights, the fifth step is to perform a structural equation modeling (SEM) analysis in which the readiness domains derived from the EFA simultaneously predict the single school outcomes construct. Prior to the SEM, all of the cycle 4 cases, as well as any cycle 3 cases lost to follow-up, will need to be removed. Although data from these cases are being used to inform the imputation process, they cannot be directly included in the final multivariate modeling phase since they have no longitudinal sampling weights from cycle 6. Hence, only the cycle 3 respondents who were followed up at cycle 6 will be used in the final SEM analysis. Note that these cases will be complete, with values imputed for variables on which missingness is due to the survey design (i.e., the readiness indicators available from cycle 4 only), as well as for variables on which missingness stems from other causes (e.g., “Don’t Know” or “Refusal” responses to the survey questions). The final SEM step will estimate the prediction weights necessary for aggregating the school readiness domains into the composite index.

It should be noted that given the large amounts of missing data that will need to be replaced prior to analysis, there exists the possibility that estimation problems could be encountered with the imputed data sets in both the EFA and/or the SEM. In such a case, we will revert to using the data from only the longitudinal component of our sample, that is, preschoolers in cycle 3 who were also assessed at cycle 6; and apply methods other than imputation (e.g., pairwise deletion, full information maximum likelihood) for accommodating the missing data in both the EFA and SEM.

Data Preparation

Data Source and Sample

The Canadian National Longitudinal Survey of Children and Youth (NLSCY) is a clustered probability sample of Canadian residential households with children aged 0-11, first conducted in 1994-1995; follow-up surveys are conducted every two years thereafter. Excluded from the sampling frame are First Nations Peoples' reserves, certain remote areas, and institutions. Information is obtained from the person most knowledgeable about that child (PMK). For ease of discussion, we will refer henceforth to the PMK as the parent. The parent completes general, parent, and child questionnaires; and standardized measures are administered to the child by trained field staff. Basic demographic information about all household members, maternal and spousal socioeconomic information, and information about the selected child are also obtained.

The present study focused on data collected in cycles 3 (1998-1999; Statistics Canada/Human Resources Development Canada, 1999), 4 (2000-2001 Statistics

Canada/Human Resources Development Canada, 2003), and 6 (2004 – 2005; Statistics Canada/Human Resources and Social Development Canada, 2006) of the NLSCY. The observed indicators of the school readiness domains were taken from cycles 3 and 4, whereas the indicators reflecting school outcomes were obtained from cycle 6. In order to address substantial information gaps regarding the school readiness indicators, a single data file was created where the cycle 3 and 4 respondents were considered to be a single sample of Canadian preschoolers aged 4 to 5.

Figure 2: Data sources used to create initial file for development of school readiness index.

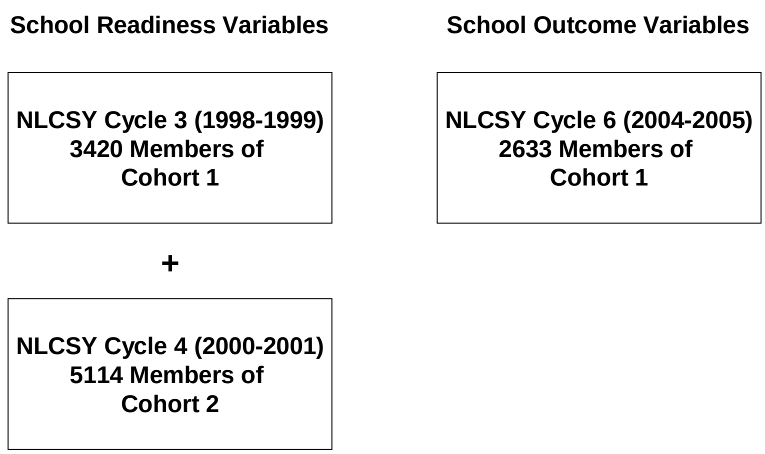


Figure 2 summarizes the data sources used to create the initial file. The group of 4- and 5-year-olds taken from cycle 3 consisted of 3,420 members of the original longitudinal cohort; specifically, those aged 0 to 1 in cycle 1 who had follow up interviews at cycle 3. At cycle 6, these same children were aged 10 to 11 and 2,633 of them were given follow-up interviews. Ideally, the analysis would have proceeded on the basis of the school readiness indicators measured at cycle 3, and the school outcomes measured at cycle 6. However, as discussed in further detail below, there was a paucity of school readiness indicators available at cycle 3. Therefore, the 3,420 cases from cycle 3 were pooled with 5,114 respondents aged 4 to 5 in cycle 4, to whom a wider variety of school readiness indicators were administered. However, the 4- and 5-year-olds from cycle 4 were not followed up at cycle 6, and therefore they do not have any data on the school outcomes. Thus the quantities of missing data in the combined data set were substantial. Table 1 displays the available sample sizes for the study variables (after the pooling of cycles 3, 4, and 6), organized in terms of the major theoretical constructs. The following section provides a more detailed description of the individual measures used.

Table 1: Available Sample Sizes for Study Variables, Based on Pooled Data From 4- and 5-year-olds (NLSCY Cycles 3, 4, and 6).

Constructs and Variables	Counts	% Total Sample
School Readiness (Cycles 3 and 4)		
Health and Physical Well Being and Motor Development		
Parent-rated health	8342	97.75
Parent-reported presence of activity limitation	8438	98.88
Parent-reported presence of chronic condition	8280	97.02
Who am I? Copying scale[a]	4047	47.42
Who am I? Symbols scale[a]	4047	47.42
Social and Emotional Development (all based on parent report)		
Hyperactivity/inattention	8250	96.67
Prosocial behaviour[b]	3077	36.06
Emotional disorder/anxiety	8275	96.97
Conduct disorder/physical aggression	8275	96.97
Indirect aggression	8024	94.02
Property Offences[c]	3286	38.50
Self-Control[a]	5002	58.61
Approaches to Learning		
Administrator observations during Peabody Picture Vocabulary Test (Revised)[b]	2705	31.70
Parent-Rated work effort[a]	4958	58.10

Parent-Rated curiosity[a]	5005	58.65
Parent-Rated novelty[a]	5011	58.72
Language Use and Communication Skills		
Peabody Picture Vocabulary Test (Revised)	7214	84.53
Parent-Reported Communication Skills[a]	4977	58.32
Cognition and General Knowledge		
Number Knowledge[a]	4438	52.00
Parent-reported independent self-care[a]	5010	58.71
School Outcomes (Cycle 6)		
Parent-Rated overall Performance	2588	30.33
Parent-Rated reading	2589	30.34
Parent-Rated writing	2586	30.30
Parent-Rated math	2589	30.34
Child-Rated overall Performance	2335	27.36
Standardized Canadian Math Achievement Test (IRT Method)	2341	27.43
Parent-Reported grade repetition	2589	30.34
Child-Reported attitude towards school	2341	27.43
Child-Reported feelings of being an outsider	2335	27.36
Parent report of discipline-related contacts to the home	2589	30.34

Note. Total sample size is 8534.
[a] Available from cycle 4 only
[b] Available at both cycles 3 and 4, but differences in item content necessitated using the cycle 3 version only
[c] Available from cycle 3 only

Measures

Health and physical well-being and motor development. Five variables were selected to represent this domain of readiness. The first was a parent-reported measure of the child's general health, common to both cycles 3 and 4 and assessed using the following question: "In general, would you say [child's name] health is: 1 = Excellent; 2 = Very good; 3 = Good; 4 = Fair; 5 = Poor." The response categories were recoded prior to analysis, so that higher scores meant better parent-rated health. Second, we used a measure of activity limitation due to a health condition. On cycle 3, this question was as follows: “Does [child’s name] have any long term conditions or health problems which prevent or limit [his/her] participation in school, at play, or in any other activity for a child of [his/her] age: 1 = yes; 2 = no.” The response categories were dummy-coded here prior to analysis (i.e., 0 = yes; 1 = no), so that a “1” meant the absence of an activity limitation. It should be noted that the activity limitation question was revised for Cycle 4; specifically, it was reworded slightly and split into three separate questions, each with a different stem concerning the types of activities in question. For example: “Does a physical condition or mental condition or health problem reduce the amount or the kind of activity [child’s name] can do: At home?” Further, the response categories were more finely graded: 1 = Yes, sometimes; 2 = Yes, often; 3 = No. Therefore, in order to pool cycle 3 and 4 children in a comparable manner on the single dummy variable, 0 = a response of “yes, often” or “yes, sometimes” to any of the three activity limitation questions on cycle 4, and 1 = a response of “no” to all of the three activity limitation questions on cycle 4.

Third, we used a dummy variable reflecting the presence of any chronic conditions (i.e., 0 = yes, 1 = no). On cycles 3 and 4, parents were asked whether the child had any of a number of long-term health conditions, which had to have lasted six months or more and been diagnosed by a health professional. Examples of conditions listed were allergies, bronchitis, and heart disease. We also ensured that our chronic condition dummy variable included asthma, which is addressed by a separate set of questions on the NLSCY.

A fourth indicator, available only from cycle 4, was the Who am I? (WAI; de Lemos & Doig, 1999) instrument, which assesses childrens’developmental level in terms of preliteracy, pre-numeracy, and fine motor skills. Specifically, summary scores on the two subscales of the WAI were used here. The *copying* scale (WAI – CS) reflects development of the ability to conceptualize a given figure, assessed using five items which range from 1 (scribble) to 4 (clear form). The *symbols* scale (WAI – SS) assesses the development of the understanding that symbols have particular meanings, also using 5 items ranging from 1 (scribble) to 4 (clear symbols). The total score on both the WAI – CS and WAI – SS can therefore range from 5 to 20.

Social and emotional development. The Child Behavior Checklist (Achenbach & Edelbrock, 1986), as modified for Canadian children, provided good coverage of social and emotional development. The scale has been previously used in two Canadian epidemiological studies: the Ontario Child Health Study (OCHS; Offord, Boyle, Fleming, Blum, & Grant, 1989) and the Montreal Longitudinal Survey (MLS; Tremblay, Pihl, Vitaro, & Dobkin, 1994). Items are scored on a 3-point scale, on which 0 indicates an absence of the behavior in question, and scores of 1 and 2 are ordered categories reflecting increasing severity and/or frequency of the behavior (e.g., 1 = Never or not true; 2 = Sometimes or somewhat true; 3 = Often or very true. Factor analyses of the instrument by methodologists at STC yielded six subscales (Statistics Canada/Human Resources Development Canada, 1999). Total scores on each of the 6 subscales were used here: *Hyperactivity/inattention* (range = 0 to 16), *Prosocial behavior* (range = 0 to 16), *Emotional disorder/anxiety* (range = 0 to 16), *Conduct disorder/physical aggression* (range = 0 to 12), *Indirect aggression* (range = 0 to 10), and *Property Offences* (range = 0 to 12). With the exception of *Prosocial behavior*, higher scores on all subscales represented more frequent or a higher level of behavior problems. Therefore, scores on the remaining subscales were reverse-coded so that a higher score represented a less frequent or a lower level of behavior problems. It should also be noted that *Property Offences* was not measured for 4- and 5-year olds in cycle 4; further, only two Prosocial behavior items were asked: “How often does [child’s name] comfort another child who is crying or upset?” and “How often does [child’s name] try to help someone who has been hurt?” Therefore, data on both *Property Offences* and *Prosocial Behaviour* were taken only from Cycle 3.

Another indicator of *Social and emotional development* a measure of self-control (Thomas, 2007), created by summing the following two items: “How often does [child’s name] keep his/her temper?” and “How often does [child’s name] show self-control?” These two items were scored on the following scale: 1 = Never; 2 = Sometimes; 3 = Often.

Approaches to learning. The construct of *Approaches to learning* was represented by a number of attitudinal and behavioural variables. One was an administrator assessment of the child’s attitude during the Peabody Picture Vocabulary Test-Revised (PPVT-R; Dunn & Dunn, 1981), which assesses children’s receptive verbal ability. Specifically, using a scale ranging from 1 (Excellent) to 5 (Poor), the administrator rated the child’s attitude toward the test, rapport with the administrator, perseverance/persistence, cooperation, and motivation/interest. Before summing the items to derive a global score, the items were reverse-coded, and then rescaled to the 0 to 4 range by subtracting 1 from the score for each item. The total score can therefore range from 0 (absence of a problem) to 24 (the highest possible score with respect to problems). We simply reversed the scoring of the final derived variable so that a higher score reflected a lower level of problems. Although this scale was used on both Cycles 3 and 4, cycle 4 used a very different set of items and response categories. Therefore, only the cycle 3 version was used, since it was these cases that would be used in the final SEM analysis to derive weights for the school readiness indicator.

Three cycle 4 items assessing the degree of effort that the child typically invests in tasks and activities (generally defined) were also used (Thomas, 2007). These items were: “How often does [child’s name] finish things he/she starts?”, “How often does [child’s name] persist with solving a problem, even when things go wrong for a while?”, and “How often does [child’s name] make an effort to do something, even if he/she doesn’t feel confident about it?” These items were all measured on a 3-point scale (1 = Never; 2 = Sometimes; 3 = Often), and were summed to create an overall “work effort” score. Other single cycle 4 items which reflected the *Approaches to learning* domain asked about the child’s typical levels of curiosity (“How often does [child’s name] ask questions or take things apart to find out how they work?”) and excitement about novelty (“How often does [child’s name] get excited about new books, toys or experiences?”). These items were also measured on a 3-point scale: 1 = Never; 2 = Sometimes; 3 = Often.

Language use and communicative skills. Our primary indicator of this domain was the Peabody Picture Vocabulary Test-Revised (PPVT-R; Dunn & Dunn, 1981), assessed at both cycles 3 and 4. The PPVT-R assesses childrens’ receptive verbal ability and is administered face-to-face by trained field staff. The PPVT-R is designed to yield a single standardized score (M = 100, SD = 15), and has been demonstrated to have strong psychometric properties (Dunn & Dunn, 1981). For example, it correlates highly with measures of intelligence, particularly verbal subscales, and is a good predictor of academic achievement.

In addition, we created another indicator from six cycle 4 items measuring childrens' interpersonal communication skills, in terms of both transmitting and receiving information (Thomas, 2007). These items were: “How often does (name of child): clearly convey his/her needs?”, “When he/she is paying attention, how often is (name of child) able to carry out a simple instruction after hearing it only once?”, “If he/she does not understand what someone has said, how often will (name of child) ask for it to be repeated or explained?”, “How often does (name of child) follow what is being talked about in a conversation, and stay on the same topic?”, “How often can (name of child) be relied on to pass simple messages from one person to another without getting the message mixed up?”, and “How often does (name of child) clearly explain about things he/she has seen or done so that you get a very good idea what happened?” These items were measured on a 3-point scale (1 = Never; 2 = Sometimes; 3 = Often), and were summed here to create a single variable reflecting interpersonal communication skills.

Cognition and general knowledge. This domain was assessed using the *Number Knowledge* direct measure (Case, 1996), assessed at cycle 4. This measure reflects childrens' intuitive number knowledge in terms of their understanding of the system of whole numbers; and yields a score representing the age equivalent for the child. A score of 0 means that the child has not reached the predimensional level; a score of 1 indicates that the child has reached the predimensional level (i.e., 4-year-old equivalent); a score of 2 means that the unidimensional level (i.e., 6-year-old equivalent) has been achieved; and a score of 3 reflects attainment of the bidimensional level (i.e., 8 year old equivalent).

Independent Self-Care. Lastly, we used an additional measure of readiness not falling under any particular domain, but which had high face validity: the child's independence in performing self-care activities (Thomas, 2007). The score on this measure was based on the following two items: "How often does [child's name] show independence with dressing?" and "How often does [child's name] show independence with washing and toileting?" These items were scored on the following 3-point scale: 1 = Never; 2 = Sometimes; 3 = Often. Scores on the two items were summed to create a single indicator of independence in performing self-care activities.

School outcomes. A number of observed attitudinal and performance measures were selected from cycle 6 to represent the school outcomes construct. First, parent-rated performance in math, reading, and writing, as well as overall performance were assessed with 4 items. Specifically, the question was, "Based on your knowledge of his/her school work, including his/her report cards, how is this child doing in the following areas at school this year?" In addition, a child-rated measure of performance was considered (Statistics Canada, 1993): "How well do you think you are doing in your school work?" Each response is scored on a 5-point scale: 1 = Very well; 2 = Well; 3 = Average; 4 = Poorly; 5 = Very poorly. Each of these items was reverse-scored prior to analysis, so that higher scores reflected better school performance ratings.

In addition, an objective measure of math performance was administered by the child's teacher, namely a shortened version of the Mathematics Computation Test (MCT) of the standardized Canadian Achievement Test, Second Edition (CAT/2; Canadian Test

Centre, 1992). More specifically, we used the STC-derived MCT score obtained via *item response theory*. To derive this score, IRT is applied to the pattern of correct answers for each combination of grade and test level (Statistics Canada, 2006). Scores increase as the child's grade level and ability increase. This score provides more precise estimates of test performance than the classically scaled scores reported for this test.

Also included was a parent-reported measure of grade repetition: “In the past two years, that is, since the end of the school year of the last interview, has this child ever repeated a grade (including kindergarten)?” This variable was coded as 0 = yes, 1 = no.

In addition to academic performance measures, it was also important to have some school outcomes measures reflecting general adjustment and adaptation to the school environment (i.e., school-related attitudes, perceptions, and behaviours; see Ladd, 2003; Ladd, Herald, & Kochel, 2006). For example, one item asked of parents was, “Since this child started school in the fall, how many times have you been contacted by phone or in writing by his school regarding his behaviour problems or challenges at school?” This item was rated on a 4-point scale: 1 = Never; 2 = Once or twice; 3 = Three or four times; 4 = Five times or more. This item was reverse-coded so that higher scores reflected less discipline-related contacts between the school and the home. We also included two child-reported measures dealing with school attitudes and perceptions. One was: “How often do you feel like an outsider (or left out of things) at your school?”, scored on a 5-point scale: 1 = All the time; 2 = Most of the time; 3 = Some of the time; 4 = Rarely; 5 = Never (World Health Organization, 1997). Another such measure was “How do you feel about school?”, scored on a 4-point scale: 1 = I like school very much; 2 = I like school quite a bit; 3 = I like school a bit; 4 = I don't like school very much (Australian Bureau of

Statistics, 1995). This item was reverse-scored prior to analysis, so that higher scores reflected more positive attitudes toward school.

Handling Missing Data: Multiple Imputation

As noted above, we used multiple imputation to replace the large volume of missing data prior to any analyses (see Table 1). As discussed previously, the cycle 3 respondents were missing on the majority of the school readiness indicators taken from cycle 4, and the cycle 4 respondents had no data for the school outcomes assessed at cycle 6. To fill in the missing values, the Markov Chain Monte Carlo (MCMC) method (Schafer, 1997) was applied, as implemented in SAS version 9.1 (SAS Institute Inc., 2003). The MCMC method is highly versatile, since the missing data patterns can be completely arbitrary (i.e., missing values can occur anywhere in the data matrix), all missing values are estimated simultaneously, and although the procedure is regression-based there is no formal distinction made between predictor and outcome variables. The MCMC procedure assumes that the data are missing at random (MAR), which means that missingness on a given variable is allowed to depend on other variables in the model, but not on the value of the variable in question (Little & Rubin, 2002). The procedure has been shown to be superior to more conventional approaches to dealing with missing data, such as listwise and pairwise deletion which assume that the data are missing completely at random (Schafer, 1997; Schafer & Olson, 2002). The various steps involved in the MCMC method are summarized below.

Step 1. A set of starting values is computed for the summary parameters (i.e., means, variances, and covariances) of the multivariate distribution of the selected variables, using an expectation-maximization (*EM*) algorithm (cf. Dempster, Laird, & Rubin, 1977). The EM process begins with a mean vector and variance-covariance matrix computed only from the available data. Next, the means, variances, and covariances are used to compute the linear regressions (intercepts and slopes) of variables with missing data on variables with complete data, within each missing data pattern (i.e., pattern of variables present and variables absent). Then, the regression equations are used to fill in the missing values within each data pattern. Since these predicted values represent the conditional mean or expectation of the variable with missing values, this is called the "expectation" (or *E*) phase. In the maximization (or *M*) phase, a maximum-likelihood estimate of the mean vector and variance-covariance matrix is obtained, based on the complete or "filled-in" raw data set constructed in the *E*-phase. The program cycles through the *E*- and *M*-phases until the mean vector and covariance matrices estimated on two successive *M*-phases differ by less than some pre-specified small value (i.e., the convergence criterion for the maximum likelihood algorithm).

Step 2. Once the *EM* mean vector and covariance matrix have been estimated, their *posterior sampling distribution* is also obtained. The posterior distribution is so-called because it is based on a *prior distribution* or "best guess", plus the actual information in the observed data. Often there is insufficient knowledge to impose a specific prior distribution on the parameter estimates (as in the present case), so a non-informative or objective *Jeffreys* prior is used as the default (cf. Schafer, 1997). Once the posterior distribution is obtained, its mode, that is, the most frequently appearing mean vector and covariance matrix, is selected as the optimal starting point for the MCMC algorithm.

Step 3. Similar to the initial *EM* steps, the MCMC technique uses the mode of the posterior distribution of the *EM* mean vector and covariance matrix (obtained in Step 2) to generate intercepts and slopes for the linear regressions of variables with missing data on variables with complete data, within each missing data pattern.

Step 4. The regression coefficients computed in Step 3 are used to fill in the missing values within each data pattern. However, unlike the *EM* algorithm, the MCMC procedure augments the predicted values used to replace the missing data. In particular, the regression equations used to impute the missing values are of the form:

$$\mathbf{X_2} = \boldsymbol{a} + \boldsymbol{b}\mathbf{X_1} + \boldsymbol{S}_{X_1X_2}\boldsymbol{E}, \qquad [3]$$

where $S_{X_1X_2}$ denotes the standard deviation of the error or residual term from the regression of X_2 on X_1, and E is a value drawn randomly from a standard normal distribution.. By adding terms such as $S_{X_1X_2}E$ to the predicted values for all cases, a correct amount of variability (i.e., random noise) is preserved within each of the imputed data sets. Consider the following example. If, for a subset of cases, complete data on variable X_1 was used to impute missing values on variable X_2 via linear regression without the augmentation described in Equation 1, the subsequent correlation between X_1 and X_2 would be 1.0 for that subset of observations. In other words, by construction X_2 would be a perfect linear function of X_1 for those observations. However, by introducing a random variation component, the association between the observed values of X_1 and the imputed values of X_2 is dampened, leading to approximately unbiased estimates of the overall correlation between the two variables (Schafer, 1997).

Step 5. After all missing values have been imputed, estimates of the summary parameters (i.e., means, variances, and covariances) are recalculated, as is their posterior sampling distribution. A random draw from the posterior distribution is then used to calculate a new set of regression coefficients for generating another round of imputed values. This random sampling of the posterior distribution ensures that the regression coefficients used to calculate the missing values are appropriately perturbed from one imputation to the next, preserving a correct amount of variability between the imputed data sets. Because the imputations are based on random draws from the posterior distribution, the MCMC technique is also referred to as *Bayesian* imputation (cf. Schafer, 1997).

In sum, the goal of the MCMC method is to impute values for the missing data that preserve the multivariate distribution of the variables, incorporating appropriate amounts of sampling error. Steps 4 and 5 are repeated several times, and the imputed data sets created on every *k*th iteration are stored for later analysis. The number of iterations is user-specified, with $k = 5$ being the default in SAS PROC MI. However, given the large quantities of missing data in this study, 20 imputed data sets were created. Schafer and Graham (2002) suggest that $k = 20$ is sufficient for most applications of multiple imputation. Further, Schafer and Olsen (1998, p. 548) provide results showing that even with a missing information fraction of 90%, parameter estimates can still have 96% efficiency with $k = 20$. It is also important to point out that the MCMC algorithm cycles through a pre-specified number of iterations before generating the first imputed data set. This process is necessary to ensure that the algorithm has converged to a stationary distribution, which can then be safely assumed to be the true posterior distribution and therefore eligible for random sampling. The iterations used to achieve stationarity of the posterior distribution are referred to as “burn-in” or “run-in” iterations (Schafer, 1997). Afterward, a pre-specified number of iterations also separates each imputed data set. In SAS PROC MI, the default number of burn-in iterations for the MCMC algorithm is 200, followed by 100 iterations between each subsequent imputed data set. To deal with the large amounts of missing data in the present study, we requested 2000 burn-in iterations, as well as 1000 between-imputation iterations.

In order to help maximize the accuracy of the imputed values, also included in the imputation process were auxiliary socio-demographic variables likely to be related to the school readiness indicators and school outcomes: the child’s gender, child’s age in months, household income in terms of total income-to-LICO (low income cut-off) ratios, and parental education. In addition, since the MCMC method assumes that the variables are multivariate normally distributed, some post-imputation adjustments were necessary to deal with the fact that some of the variables used here were binary (e.g., presence of chronic conditions, activity limitation, grade repetition) or ordered-categorical (e.g., parent-rated health, attitudes toward school, parent- and self-rated school performance). In particular, any imputed non-integer values for such variables were rounded to discrete values. For binary variables, imputed values between 0 and 1 had to exceed .6 to be assigned a value of 1. Any out-of-range imputed values were accordingly assigned the lowest or highest values on the scales for the variables in question. Imputed values on continuous variables were simply rounded to the nearest integer. These adjustments have been shown to still yield unbiased estimates, and the MCMC procedure is quite robust to non-normality (Schafer & Graham, 2002).

It should be noted that there is one other missing data technique which is frequently recommended over multiple imputation: *full information maximum likelihood* (FIML) estimation. Under the FIML technique, all cases are included in the estimation of model parameters, provided that they have at least one data point across all of the observed variables being analyzed. Missing values are not actually replaced, but no cases with partial data are wasted. FIML estimation has been shown to outperform both listwise and pairwise deletion, as well as multiple imputation in the handling of missing values; and is regarded as the missing data method of choice, particularly for SEM (Allison, 2003; Arbuckle, 1996; Enders & Bandalos, 2001; Peters & Enders, 2002; Wothke, 2000). Unfortunately, since certain variables used in the present study were missing by design, FIML was unable to be used with the pooled data set. The reason is that for each pair of variables in the analysis, a non-zero proportion of covariance coverage is required in order for FIML to work (Muthén & Muthén, 2005). For example, certain school readiness measures were only administered to the 4- and 5-year-olds at cycle 4, and these children have no follow-up data on the school outcomes at cycle 6. Therefore, in the initial pooled data set, there will be zero coverage of the associations between the cycle 4 readiness indicators and the cycle 6 school outcomes. However, we reserve FIML as an option for analyzing just the cycle 3 and cycle 6 data, if the current imputation strategy fails.

Estimating the School Readiness Model

Exploratory Factor Analysis (EFA)

Once the 20 imputed data sets were generated, the indicators of the school readiness domains were subjected to exploratory factor analysis (EFA) to determine their optimal sub-groupings. Since the school readiness outcomes were not part of this analysis, we could use the data for both the cycle 3 and cycle 4 children to examine the dimensionality of the indicators (N = 8534). The EFA was carried out using the M*plus* program (version 4.2; Muthén & Muthén, 2005). Given that many of the observed variables had binary or ordinal response categories, a Robust Weighted Least Squares (RWLS) estimation procedure was used (Flora & Curran, 2004; Muthén, du Toit, & Spisic, in press). To determine the optimal number of factors for extraction, we applied the Kaiser criterion: the number of eigenvalues ≥ 1.0 (see Comrey & Lee, 1992). Eigenvalues are repackaged information from the correlation matrix of the observed items, and each represents the amount of variance explained by a factor; factors with eigenvalues ≥ 1.0 are considered to represent meaningful amounts of explained variation. When analyzing imputed data, only the eigenvalues based on the first imputed data set are presented, and these are used to guide further factor extraction (Muthén & Muthén, 2005). To enhance interpretability of the factor solutions, promax rotation was applied, which allows factors to intercorrelate. It seemed reasonable that any obtained factors would be interrelated, given that they fall under the broad umbrella of school readiness. Factor loadings ≥ 0.30 were considered to indicate reliable saturation of observed items by latent factors. In the present analysis, the estimates of the loadings are simply the average across all 20 imputed data sets (Rubin, 1987).

Since no standard errors are conventionally produced in an EFA, there was no need to adjust for the complex sampling design of the NLSCY (i.e., stratification and clustering). However, cross-sectional sampling weights were applied to take certain sampling features into account including unequal probabilities of selection, non-response (person and household level), and an adjustment making the age and sex distributions of the sample correspond to the age and sex distributions of the Canadian population (Statistics Canada/Human Resources Development Canada, 1996). Applying these weight helps ensure unbiased estimates of the factor model parameters (i.e., factor loadings and intercorrelations; Asparouhov, 2006; Kaplan & Ferguson, 1999). In this, and all subsequent analyses, we used the natural logarithm of the scores for the PPVT-R and the IRT-based Math Scores, in order to reduce their variances (which were substantially higher than those of all other variables in the analysis) and thereby avoid potential computational problems (Muthén & Muthén, 2005).

A variety of indices were used to assess the global fit of the factor model to the imputed data sets. To statistically test the difference between the model and the data, we used the chi-square (χ^2) statistic, aggregated across all 20 model runs. However, χ^2 is correlated with sample size if the model is not perfectly correct (Browne & Cudeck, 1993). Given that most statistical models are at best approximations to reality rather than exact matches, it is usually not reasonable to demand a nonsignificant χ^2 test (McDonald & Marsh, 1990). Therefore, we supplemented the χ^2 test with two widely used approximate fit indexes that are relatively unaffected by sample size: the Root Mean Squared Error of Approximation (RMSEA; Steiger, 2000) and the Standardized Root Mean Squared Residual (SRMR; Jöreskog & Sörbom, 1996). In terms of criterion values for these indexes, a RMSEA ≤ .06 and a SRMR ≤ .08 are indications of a good-fitting model (Hu & Bentler, 1999). In addition to the above statistical criteria, parsimony and ease of interpretability were used to select the optimal factor solution.

Table 2: Eigenvalues ≥ 1.0 in the factor solution for the first imputed data set

Factor	**Eigenvalue**
1	**3.91**
2	**2.22**
3	**1.83**
4	**1.63**
5	**1.42**
6	**1.18**
7	**1.01**
***Note.* Sample size = 8534**	

Based on the first imputed data set, a total of seven important eigenvalues were found in the correlation matrix of the observed variables, and are displayed in Table 2. However, due to computational problems, the Mplus program was unable to extract more than two of these identified factors. Specifically, extraction of more than two factors resulted in a number of negative error variances for the observed variables (i.e., variance not explained by the factor solution), which are inadmissible parameter estimates known as Heywood cases (Chen, Bollen, Paxton, Curran, & Kirby, 2001).

Table 3: 2- factor solution across 20 imputed data sets, promax rotation

Variables	**Loadings**	
	Factor I	**Factor II**
Parent-Rated Health	0.26	0.07
Presence of Activity Limitation	0.21	0.18
Presence of Chronic Condition	0.17	0.02
Independent Self-Care	0.24	0.15
Hyperactivity/inattention	**0.62**	0.10
Emotional disorder/anxiety	**0.48**	-0.10
Conduct disorder/physical aggression	**0.66**	-0.11
Indirect aggression	**0.45**	-0.16
Property offences	**0.67**	-0.10
Prosocial behaviour	0.05	0.13
Self-control	0.53	0.05
PPVT-R	0.04	**0.40**
PPVT-R Administrator observations	0.04	**0.34**
Work Effort	**0.44**	0.17
Curiosity	0.14	0.04
Novelty	0.14	0.11
Communication Skill	**0.37**	0.22
Number Knowledge	-0.06	**0.73**
Who am I? Copying Scale	0.01	**0.69**
Who am I? Symbols Scale	-0.01	**0.62**
Note. Sample size = 8534. Factor loadings ≥ 0.30 are bolded.		

The two-factor solution differed from the data according to the χ^2 test (χ^2 [N = 8534, df = 69] = 2323.08, p < .001), yet fit well according to the RMSEA (.06) and SRMR.

The factor loadings are displayed in Table 3. The first factor appears to consist of primarily behavioural and attitudinal items, while the second is reflected primarily by the mental ability tests. However, given (1) the lack of salient loadings for many of the key variables (e.g., none of the health-related variables loaded on either factor) and (2) the technical problems encountered when trying to extract more than two of the seven important factors initially located, it was concluded that the imputed data were not sufficiently trustworthy for the analytical phase of the current study. Therefore, the second EFA attempt focused on only the original, unimputed cycle 3 data for the school readiness predictors (*N* = 3420). The indicators of school readiness were thus reduced to the following: parent-rated health, activity limitation, presence of chronic conditions, PPVT-R, administrator-assessed attitude during PPVT-R, hyperactivity/inattention, prosocial behavior, emotional disorder/anxiety, conduct disorder/physical aggression, indirect aggression, and property offences. Pairwise deletion was used to accommodate the missing data. Specifically, for each pair of observed variables, each correlation in the matrix to be factor-analyzed was calculated using respondents with complete data on both variables. The only cases completely excluded from the analysis are those which were missing data for all items. While FIML estimation would have been preferable to pairwise deletion here, FIML is at present only available in Mplus for EFA with all continuous variables (Muthén & Muthén, 2005).

There were 48 cases excluded from the EFA due to having missing data on all variables (*N* = 3372). A total of three important eigenvalues were found in the correlation matrix, and accordingly a 3-factor solution was subjected to promax rotation. However, the prosocial behaviour subscale failed to load reliably on any of the factors. Therefore, this variable was dropped and the EFA was carried out again. Three substantial eigenvalues still emerged, and promax rotation produced an interpretable 3-factor solution (displayed in Table 4).

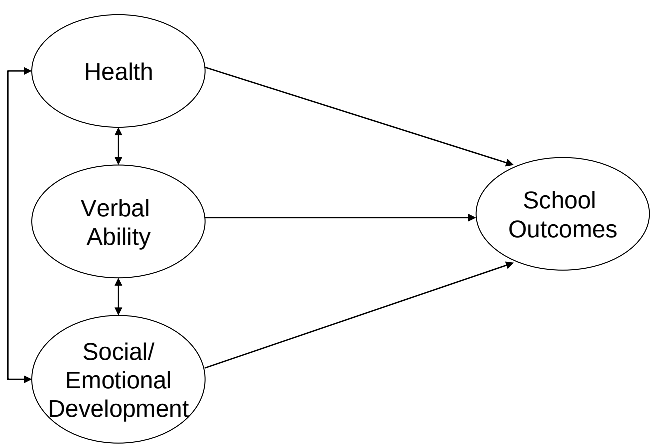
Table 4: 3-factor solution for school readiness indicators from NLSCY cycle 3

Observed Variables	**Factors**		
	Health	**Social/Emotional Development**	**Verbal Ability**
Parent-Rated Health	**0.41**	0.18	-0.03
Activity Limitation	**0.72**	-0.02	0.29
Chronic Condition	**0.91**	-0.06	-0.11
Hyperactivity/Inattention	-0.01	**0.63**	0.09
Emotional Disorder/Anxiety	0.07	**0.47**	-0.02
Conduct Disorder/Physical Aggression	-0.00	**0.73**	-0.02
Indirect Aggression	-0.09	**0.46**	-0.04
Property Offences	-0.06	**0.72**	-0.01
PPVT-R	-0.00	-0.00	**0.50**
PPVT-R –Administrator Assessment	-0.03	0.02	**0.72**
Note: Sample size is 3372. Factor loadings ≥ 0.30 are bolded.			

While the model differed significantly from the data according to the χ^2 test (χ^2 [*df* = 11, *N* = 3372] = 27.938, $p < .001$), it demonstrated a good fit according to the approximate fit indexes (RMSEA = .021; SRMR = .026). The pattern of salient loadings (≥ .30) suggests the following three factors: **Health** (Parent-rated health, activity limitation, and the presence of one or more chronic conditions), **Verbal Ability** (standardized score for PPVT-R, administrator-assessed attitude toward PPVT-R), and **Social/Emotional Development** (hyperactivity/inattention, emotional disorder/anxiety, conduct disorder/physical aggression, indirect aggression, and property offences). All complex or cross-loadings were < .30, supporting a clean factor solution and straightforward interpretation.

Because of the reduction in the initial number of observed indicators of school readiness (available from the combined data for NLSCY cycles 3 and 4), the labels given to the three factors revealed here for the cycle 3 data are not perfectly matched with any of the readiness domains initially postulated. However, the 3-factor solution is conceptually coherent, and the factors are reasonably empirically distinct. **Health** was modestly correlated with both **Social/Emotional Development** (r = 0.279) and **Verbal Ability** (r = .273), and **Social/Emotional Development** and **Verbal Ability** were also mildly yet significantly correlated at r = 0.131. This model therefore served as the basis for specifying the SEM used to predict the cycle 6 school outcomes from the cycle 3 readiness domains. Accordingly, Figure 3 shows a revised conceptual model, based on the EFA results from the cycle 3 data.

Figure 3: Revised EFA-Based Conceptual Model Depicting Three School Readiness Domains as Simultaneous Predictors of School Outcomes. (Measured variables reflecting the domains are not shown here but rather described in the text).



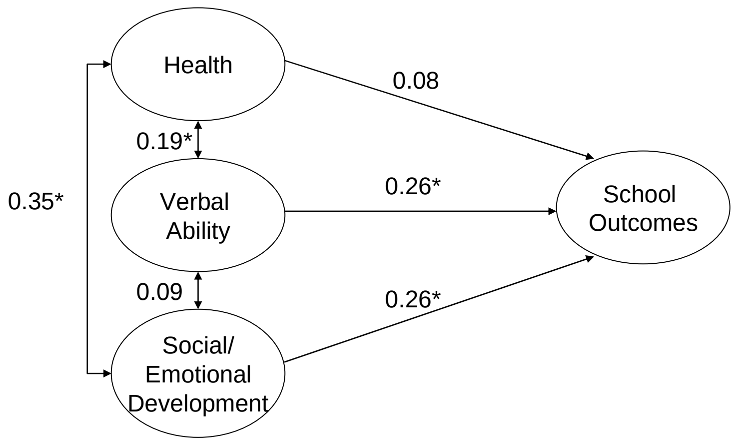
Structural Equation Modeling (SEM)

The model derived from the EFA and depicted in Figure 3 was evaluated using Structural Equation Modeling (SEM), as implemented in the M*plus* software package (Version 4.2; Muthén & Muthén, 2005). Specifically, a school outcomes construct was defined by the following school attitude, achievement, and performance variables: parent-rated math, writing, reading, and overall performance, child-rated overall performance, grade repetition, number of discipline calls to the home, attitude toward school, and the extent to which the child feels like an "outsider" in school. This outcomes factor was regressed on the three factors identified in the EFA of the cycle 3 school readiness indicators. In the SEM, each readiness factor was defined by exactly the same indicators that showed salient loadings on that factor from the EFA. The three factors were still allowed to intercorrelate, but all cross-loadings identified in the EFA were dropped from the SEM, since they were all < .30. The factor loadings of binary and ordered-categorical indicators were specified as probit coefficients, and factor loadings of continuous indicators were specified as linear coefficients. Directional relations among the latent variables were also specified as linear coefficients, and bidirectional relations were specified as Pearson correlations. The variance of each factor was fixed at 1 for parameter identification purposes, and this strategy also ensured that each factor was standardized, allowing for interpretation of the factor scores as normalized (i.e., *z*) scores (Muthén & Muthén, 2005).

Longitudinal sampling weights were applied in order to get unbiased estimates of the model parameters (Asparouhov, 2006; Kaplan & Ferguson, 1999). Information on stratification and clustering was used to adjust the standard errors for the nonindependence of observations induced by the complex sampling design (Asparouhov, 2006). Since n = 787 of the cycle 3 cases were lost to follow-up and therefore had no cycle 6 data on school outcomes, the available sample size for the SEM analysis was n = 2633.

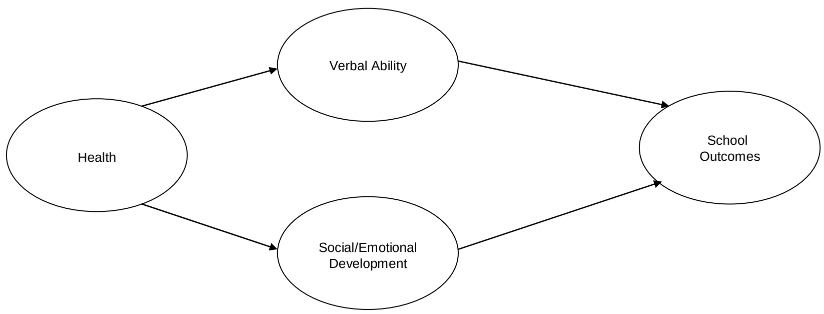
In order to accommodate missing data in the SEM with a combination of continuous and categorical variables, a FIML estimator was used, with a numerical integration algorithm (Muthén & Muthén, 2005). Since this procedure requires the raw data rather than just summary statistics (e.g., means and covariances or correlations), no conventional SEM fit tests (e.g., χ^2) or indexes (e.g., RMSEA, RMSR) are produced. Without summary statistics, there is no logical or natural “saturated” (i.e., perfectly fitting) model with which to compare the fit of the restrictive theoretical model (cf. Bollen, 1989; Hayduk, 1987). However, standard errors are still produced and can be used to evaluate individual parameter estimates. The standard errors were also adjusted for any non-normality in the observed data (Muthén & Muthén, 2005; Yuan & Bentler, 2000).

Figure 4: Results for initial SEM, in standardized form; N = 2633
Note 1: Significant coefficients are marked with a *
Note 2: All factor loadings (not shown) are significant at the .01 level



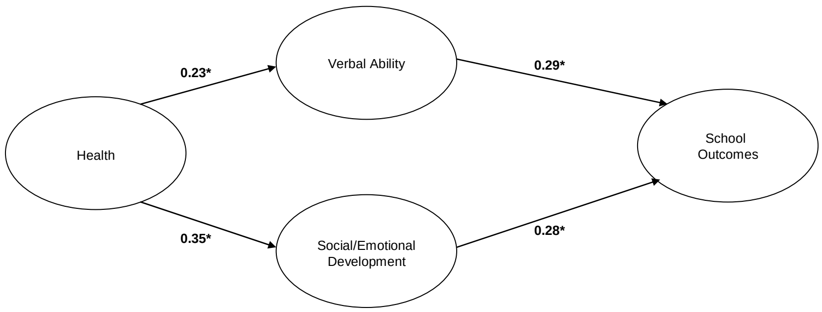
The results are displayed in Figure 4, in standardized form. All factor loadings were significant at the .01 level (not shown). Both Verbal Ability and Social/Emotional Development significantly predicted the School Outcomes Construct, but Health did not. However, Health was still correlated significantly with both Verbal Ability and Social/Emotional Development. Therefore, an alternative model specification was to consider Health as a background variable affecting both Verbal Ability and Social/Emotional Development (see Figure 5), which in turn influenced school outcomes. This specification allowed Health to still play a role in the readiness framework; specifically, the effects of both Verbal Ability and Social/Emotional Development on School Outcomes could be adjusted for Health.

Figure 5: Revised SEM; Health is now specified as a background factor influencing both Verbal Ability and Social/Emotional Development.



The results for the final model are displayed in Figure 6, in standardized form. All factor loadings were significant at the .01 level, as were the directional paths among the constructs. Health significantly predicted both Verbal Ability ($\beta = 0.23$, $p < .01$) and Social/Emotional Development ($\beta = 0.35$, $p < .01$). In turn, Verbal Ability significantly predicted School Outcomes ($\beta = 0.29$, $p < .01$), as did Social/Emotional Development ($\beta = 0.28$, $p < .01$). In total, the model explained 18% of the variance in school outcomes.

Figure 6: Results for final SEM, in standardized form; N = 2633
Note 1: Significant coefficients are marked with a *
Note 2: All factor loadings (not shown) are significant at the .01 level



Based on the results shown in Figure 6, the prediction equation for generating a school readiness score for a given case *i* is (using standardized coefficients and thereby eliminating the intercept term):

School Readiness$_i$ = 0.29***Verbal Ability**$_i$ + 0.28***Social/Emotional Development**$_i$,

[4]

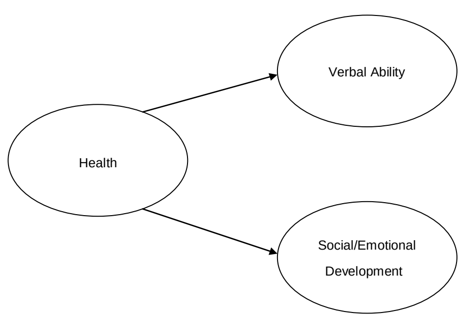
where the factor scores for Verbal Ability and Social/Emotional Developmenthave been adjusted for the effects of child Health. After deriving the statistical function for computing school readiness, casewise scores on a school readiness index can be conveniently computed from domain scores, without the need for longitudinal data on school outcomes. A prediction equation considerably reduces the expense of obtaining data on school outcomes, allowing researchers and policy makers to regularly assess, monitor, and report on school readiness as more nationally representative data on the domains are produced. However, it should be noted that the creation of the initial function for producing the index necessitates longitudinal data.

PART II: Computation of School Readiness Index for Canadian Preschoolers and Time Series Analysis

Overview

After deriving a prediction equation for computing individual-level school readiness scores, the next component of this study focused on generating a time series of provincial school readiness estimates, based on the four most recent cycles of the NLSCY; specifically, for the years 1998, 2000, 2002, and 2004. This analysis was also a multiple-step endeavour, which is summarized here and presented in more detail below.

Figure 7: Model used to generate factor scores for computing school readiness index



First, for all cycles of NLSCY data used here to produce the time series, factor scores needed to be generated on the two readiness domains of Verbal Ability and Behaviour, as well as on Health. This was accomplished by estimating a simplified version of the model shown in Figure 5; in particular, since school outcomes are no longer required, the final model examined in the SEM reduces to the form shown in Figure 7. Using the exact same estimation procedure as in the initial SEM used to derive the index, the model is applied to the observed indicators of these three constructs from NLSCY cycles 3 (i.e., the initial calibration sample where the prediction equation was developed; Statistics Canada/Human Resources Development Canada, 1999), 4 (Statistics Canada/Human Resources Development Canada, 2003), 5 (Statistics Canada/Social Development Canada, 2005), and 6 (Statistics Canada/Human Resources and Social Development Canada, 2006) for preschoolers (i.e., 4- and 5-year-olds). Estimating the model produced standardized factor scores for each construct, for all four NLSCY cycles examined here. Further, in accordance with the model in Figure 7, Health is specified as a predictor of both Verbal Ability and Social/Emotional Development, thereby adjusting the factor scores on the latter two domains for the effects of Health. Second, equation 4 was applied to the health-adjusted factor scores for Verbal Ability and Social/Emotional Development for, in order to compute school readiness scores:

$$\textbf{School Readiness}_i = 0.29*\textbf{Verbal Ability}_i + 0.28*\text{Social/Emotional Development}_i, \quad [4]$$

Third, the school readiness scores were converted to normal curve equivalents (NCEs), for both descriptive and comparative purposes at the provincial level.

Estimating Factor Scores for Readiness Domains

In order to estimate factor scores on the readiness domains, the model shown in Figure 7 needed to be estimated at each NLSCY cycle (3, 4, 5, and 6) for 4- and 5- year-olds. Table 5 shows the sample sizes available at each cycle, broken down by key socio-demographic variables.

Table 5: Sample sizes for 4- and 5-year-olds at NLSCY cycles 3 – 6, broken down by key socio-demographic variables.

Socio-demographics	NLSCY Cycle			
	Cycle 3 (1998-1999) N = 2633	Cycle 4 (2000-2001) N = 5063	Cycle 5 (2002-2003) N = 6148	Cycle 6 (2004-2005) N = 3325
Child Sex				
Male	1315	2576	3058	1719
Female	1318	2487	3090	1606
Missing	--	--	--	--
PMK Education				
< High School	249	582	709	327
Secondary school graduation	410	930	1410	747
Beyond high school	734	1074	964	426
Post-secondary graduation (inc. trade)	1238	2394	2965	1781
Missing	2	83	100	44
Ratio of Total Family Income to LICO				
< 0.75	246	461	559	326
≥ 0.75 and < 0.9	101	218	267	114
≥ 0.9 and < 1.0	81	187	168	95
≥ 1.0 and <1.1	99	163	215	101
≥ 1.1 and <1.2	157	291	376	174
≥1.25	1898	3700	4562	2515
Missing	51	43	1	--
Province of Residence				
Newfoundland	135	415	481	221
Prince Edward Island	72	253	224	190
Nova Scotia	190	457	494	242
New Brunswick	154	420	465	247
Quebec	532	781	1035	496
Ontario	662	915	1390	830
Manitoba	221	425	478	270
Saskatchewan	208	381	461	249
Alberta	259	512	614	289
British Columbia	200	504	506	291
Missing	--	--	--	--

Note. Total sample sizes are based on preschoolers eligible for SEM analysis (i.e., at least one data point across all observed indicators of the school readiness domains), and therefore for whom school readiness index scores could be generated.

Ideally, the models fitted at each NLSCY cycle would have been identical in terms of the observed indicators that were used in the initial SEM to develop the prediction equation for school readiness: Health (Parent-rated health, activity limitation, and the presence of one or more chronic conditions), Verbal Ability (standardized score for PPVT-R, administrator-assessed attitude toward PPVT-R), and Social/Emotional Development (hyperactivity/inattention, emotional disorder/anxiety, conduct disorder/physical aggression, indirect aggression, and property offences). However, certain indicators were not measured at all cycles. Specifically, the administrator-assessed attitude toward the PPVT-R was not measured at cycles 5 and 6. Therefore, only the score on the PPVT-R itself was used as an indicator of Verbal Ability at these two cycles. Further, property offences were not measured in cycles 4 through 6, and thus this subscale could not be used as an indicator of Social/Emotional Development for these cycles. The remaining observed indicators were all available, however, allowing the same three readiness constructs to still be modeled at each cycle.

Technical details on the estimation of factor scores within the M*plus* program are provided in Muthén (1998-2004, pp. 47-48). Essentially, the factor scores are held to be the “true” or error-free scores for the respondents on the latent constructs – here, Health, Behaviour, and Verbal Ability – estimated with respect to the factor model and the pattern of scores on the observed variables. Once the factor scores were estimated, equation 4 was applied in order to produce casewise school readiness scores for all 4- and 5-year-olds in NLSCY cycles 3 through 6.

Computing Normal Curve Equivalents for the School Readiness Index

After generating the individual-level school readiness scores, the final step was to convert these into a suitable form for descriptive and comparative purposes, within a time series analysis. Sine the school readiness index is essentially a weighted function of scale-free latent variables, its true theoretical range cannot be determined, and thus no conversion technique would yield a meaningful absolute score. Therefore, we searched for a conversion technique that would allow for interpretation of school readiness levels in a relative sense.

For this purpose, we selected the normal curve equivalent (NCE), a statistic developed for the United States Department of Education by the RMC Research Corporation (Mertler, 2002). NCEs are based on percentile ranks and essentially convert percentile ranks into an equal-interval scale. Therefore, unlike percentile ranks, NCEs can be used to create summary statistics (e.g., means). In particular, calculation of the NCE involves (1) assigning percentile ranks (i.e., 1 to 99%) to the raw scores, (2) assigning a *z*-score to each case based on its percentile rank, and (3), rescaling the z-scores so as to preserve the 1 to 99% range of the percentile rank. Formally, the NCE is computed here as:

$$NCE_i = (\Phi^{-1}(PR_i/100))*21.06 + 50 \qquad [5],$$

where Φ^{-1} is the inverse of the standard normal cumulative distribution function and PR_i is the percentile rank for respondent *i*'s predicted level of school readiness. The NCE is therefore assigned a mean of 50%, with a standard deviation unit being represented by a 21.06% increase or decrease from the mean. The NCE has a range of 1 to 99% and approximates percentile rank norms.

In the current time series analysis (1998, 2000, 2002, and 2004) based on NLSCY data, we examined how provincial levels of school readiness, expressed as within-province NCE means, were distributed around the imposed national mean of 50%. Results are presented in Table 6.

Table 6: Mean provincial estimates of school readiness for 1998, 2000, 2002, and 2004, based on normal curve equivalents

Province	Year							
	1998		2000		2002		2004	
	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
Newfoundland	45.13*	41.95,48.31	53.31*	51.29,55.35	52.05*	50.31,53.78	53.19*	50.62,55.75
Prince Edward Island	47.75	43.97,51.52	50.67	48.06,53.28	51.48	49.04,53.92	51.69	48.93,54.45
Nova Scotia	44.04*	41.68,46.40	48.68	46.89,50.48	48.47	46.68,50.26	50.06	47.37,52.75
New Brunswick	43.54*	40.55,46.52	49.04	46.81,51.27	52.46*	50.71,54.21	51.69	49.18,54.19
Quebec	48.66	46.74,50.57	44.42*	42.90,45.94	45.34*	44.03,46.64	43.95*	42.02,45.88
Ontario	51.72*	50.21,53.23	47.71*	46.38,49.04	51.24*	50.10,52.39	49.98	48.53,51.42
Manitoba	50.24	46.84,53.63	47.50*	45.71,49.29	51.83	49.98,53.69	51.06	48.54,53.67
Saskatchewan	51.60	48.58,54.62	45.50*	43.54,47.47	47.22*	45.30,49.14	48.03	45.38,50.69
Alberta	53.59*	50.92,56.25	50.08	48.52,51.64	48.04*	46.31,49.77	50.04	47.66,52.42
British Columbia	45.48*	42.66,48.29	48.49	46.62,50.36	48.39	46.56,50.23	51.03	48.39,53.67

***Note*. Data Source: National Longitudinal Survey of Children and Youth.**
*** Significantly different from the fixed national mean of 50%.**

As shown in Table 6, there are some temporal fluctuations in the provincial school readiness estimates in relation to the fixed national mean. And for the most part the changes are not strictly monotonic (i.e., steadily or smoothly increasing or decreasing over time). The estimates for which the confidence intervals do not include the fixed national mean (50%) indicate significant discrepancies from the national mean. Some particular patterns are noteworthy. For example, in 1998, Newfoundland, Nova Scotia, and New Brunswick, and British Columbia were all significantly lower than the national mean. Newfoundland moved to being significantly greater than the national mean for all of the remaining years in the series. Nova Scotia shifted to the level of the national mean in 2000 through 2004, as did British Columbia. New Brunswick was equivalent to the national mean in 2000, greater in 2002, and then returned to the level of the national mean in 2004.

Prince Edward Island was not significantly different from the national mean in any of the years examined. Ontario showed the highest degree of fluctuation, being significantly higher than the national mean in 1998, lower in 2000, higher in 2002, and then falling to the level of the national mean in 2004. Manitoba was significantly lower than the national mean in both 1998 and 2000, but not significantly from it in both 2002 and 2004. Saskatchewan was not significantly different than the national mean in 1998, was lower in both 2000 and 2002, but was at the level of the national mean again in 2004. Alberta was significantly higher than the national mean in 1998, not reliably different from the national mean in 2000, lower in 2002, and again not significantly different in 2004 Quebec was not significantly different from the national mean in 1998, but was significantly lower for the remainder of the time series.

Discussion

With data on school readiness domains and school outcomes from a nationally representative longitudinal sample of Canadian preschoolers aged 4 and 5, this study used a combination of exploratory factor analysis (EFA) and structural equation modeling (SEM) to derive a weighted statistical function for computing individual-level school readiness scores. Once the function is estimated based on longitudinal data, it provides a convenient means of estimating levels of school readiness from cross-sectional data, without having to collect longitudinal information on school outcomes for the same children. This function was applied to cross-sectional data for 4- and 5-year-olds from four cycles of the NLSCY in order to generate school readiness scores, which were then converted into normal curve equivalents (NCEs) and aggregated at the provincial level for a time series analysis; specifically, for the years 1998, 2000, 2002, and 2004. The results revealed interesting temporal changes in mean provincial school readiness levels in relation to a fixed national mean. It seems possible that the observed fluctuations are associated with policy and program changes, but addressing this issue would require additional “drill-down” work, of both a quantitative and qualitative nature. While examining the underlying reasons for these differences is beyond the scope of the present study, this is a promising avenue for future research.

There are a number of limitations of the present study that need to be considered when interpreting and using the findings. First and foremost is the fact that in the final analysis, there was fairly narrow empirical coverage of the five school readiness domains originally identified (see Figure 1), which were: (i) Health and Physical Well-Being and Motor Development, (ii) Social and Emotional Development, (iii) Approaches to Learning, (iv) Language and Communication Skills, and (v) Cognition and General Knowledge. Due to data gaps, coverage of the readiness domains for our longitudinal sample of preschoolers was quite sparse. To deal with this problem, we did attempt some methodological innovations, namely pooling our longitudinal sample with a cross-sectional sample in which more readiness domains were assessed, followed by multiple imputation to fill in the missing values. However, due to the vast amount of missing information to be replaced, technical difficulties emerged in the statistical modeling of the imputed data sets, and consequently we needed to revert to using the longitudinal sample only, limited by the assessment of fewer school readiness constructs. The analysis identified three domains of readiness that were important for predicting readiness scores

– Verbal Ability, Social/Emotional Development, and Health as a background adjustment factor.

Second, as noted previously, certain indicators were not measured at all cycles, and so the predicted values on school readiness were not based on exactly the same inputs. Specifically, the administrator-assessed attitude toward the PPVT-R was not measured at cycles 5 and 6. Therefore, only the score on the PPVT-R itself was used as an indicator of Verbal Ability at these two cycles. Further, property offences were not measured in cycles 4 through 6, and thus this subscale could not be used as an indicator of Behaviour in these three cycles. While the three constructs (Health, Social/Emotional Development, and Verbal Ability) were still able to be assessed, consistency across time in measuring these concepts, with identical wording and numbers of items, is desirable.

Third, since the theoretical range of the school readiness index is unknown, we were unable to provide meaningful absolute scores or criterion values against which to assess progress. The NCEs provided a relative assessment of provincial school readiness levels with respect to a fixed national mean, and therefore gives some information in terms of inequalities; however they do not give an idea of how far these levels are from a scientifically established “standard” for school readiness.

Despite these shortcomings, we feel that the statistical methodology used in the current study is sound and provides a solid template for constructing national indexes from survey data. A crucial next step in addressing the information gaps in the present school readiness index is the exploration of other longitudinal, nationally representative data sets containing a broader array of variables on school readiness and school outcomes for a given cohort. Analysis of such data sets with the statistical techniques used here has the potential to validate the methods established here, as well as greatly improve both the breadth and depth of the information encapsulated in the school readiness index, in order to produce a more reliable means of assessing school readiness for the purposes of both research and policy-making.

References

Achenbach, T. M., & Edelbrock, C. S. (1986). *Manual for the Child Behavior Checklist*. Burlington: University of Vermont.

Allison, P.D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology, 112(4)*, 545-557.

Arbuckle, J.L. (1996). Full information estimation in the presence of missing data. In G.A. Marcoulides and R.E. Schumacker (Eds.), *Advanced structural equation: Issues and techniques* (pp. 243-277). Mahwah, NJ: Erlbaum.

Asparouhov, T. (2006). Sampling weights in latent variable modeling. *Structural Equation Modeling, 12*, 411-434.

Australian Bureau of Statistics. (1995). *Western Australia: Child Health Survey: Education and Health, 1995*. Canberra: Australian Bureau of Statistics. Accessed on January 17, 2008 at: http://www.abs.gov.au/AUSSTATS/abs@.nsf/productsbyCatalogue/5EFAE6CCFACE5DFFCA25722E001A3725?OpenDocument.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Britto, P. R., Kohen, D. E., Engle, P., Bickel, S., Ulkuer, N. & Dawson, C. (2006). *Review and Analyses of International Assessment Practices*. Report prepared for UNICEF, NY, NY.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Canadian Test Centre. 1992. *CAT/2: Canadian Achievement Tests Second Edition:*

Technical Bulletin. Markham, Ontario, Canada.

Case, R. (1996). Reconceptualizing the nature of children’s conceptual structures and their development in middle childhood. In R. Case & Y. Okamoto (Eds.), The role of central conceptual structures in the development of children's thought, *Monographs of the Society for Research in Child Development*, 61 (1-2), Serial No. 246, 1–26.

Chen, F., Bollen, K.A., Paxton, P., Curran, P., & Kirby, J. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods & Research*, *29*, 468-508.

de Lemos, M., & Doig, B. (1999). *Who Am I?: Developmental Assessment*. Melbourne, Australia: ACER.

Comrey, A.L., & Lee, H.B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Dempster, A.P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, *B*, *39*, 1-38.

Dunn, L. M. & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test-Revised*. Circle Pines, MN: American Guidance Service.

Enders, C.K., & Bandalos, D.L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. Structural Equation Modeling, 8(3), 430-457.

Flora, D.B., & Curran, P.J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466-491.

Hayduk, L.A. (1987). *Structural equation modeling with LISREL: Essentials and advances*. Baltimore, MD: Johns Hopkins University Press.

Hu, L-T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Jöreskog, K.G., & Sörbom, D. (1996). *LISREL 8 user's reference guide*. Chicago: Scientific Software International, Inc.

Kaplan, D., & Ferguson, A.J. (1999). On the utilization of sample weights in latent variable models. *Structural Equation Modeling*, *6*, 305-321.

Kohen, D. (2006). *School Readiness: Domains and Influences*. Report prepared for the Canadian Council on Learning, Aug. 2006. Ottawa, ON.

Kohen, D. & Britto, P. (2006). *Theories of School Readiness*. Background paper prepared for the Canadian Council on Learning, Aug.2006. Ottawa, ON.

Ladd, G.W. (2003). *School Transitions/School Readiness: An Outcome of Early Childhood Development*. Encyclopedia on Early Childhood Development. Centre of Excellence for Early Childhood Development, Arizona State University, USA. Available online at http://www.excellence-earlychildhood.ca/documents/LaddANGxp.pdf.

Ladd, G.W., Herald, S.L., & Kochel, K.P. (2006). School readiness: Are there Are there social prerequisites? Early Education and Development, *17(1)*, 115–150.

Little, R.J.A., & Rubin, D.B. (2002). *Statistical Analysis with Missing Data* (2nd ed.). New York: Wiley.

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness-of-fit. *Psychological Bulletin, 107*, 247-255.

Mertler, C. A. (2002). *Using standardized test data to guide instruction and intervention.* College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

Muthén, B.O. (1998-2004). *Mplus Technical Appendices*. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2005). *Mplus Version 4.2 user's guide*. Los Angeles, CA: Muthén & Muthén.

Muthén, B. O., du Toit, S. H. C., & Spisic, D. (in press). Robust inference using weighted-least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. *Psychometrika*.

Offord, D. R., Boyle, M. H., Fleming, J. E., Blum, H. M., & Grant, N.I. (1989).Ontario Child Health Study: Summary of selected results. *Canadian Journal of Psychiatry*, *34*, 483-91.

Peters, C.L.O., & Enders, C.K. (2002). A primer for the estimation of structural equation models in the presence of missing data: Maximum likelihood algorithms. *Journal of Targeting, Measurement and Analysis for Marketing*, *11(1),* 81-95.

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

SAS Institute Inc. (2003). *SAS Version 9.1 User's Guide*. Cary, NC: SAS Institute Inc.

Schafer, J.L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.

Schafer, J.L., & Graham, J.W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7(2)*, 147-177.

Schafer, J.L., & Olsen, M.K. (1998). Multiple imputation for multivariate missing data problems: A data analyst's perspective. *Multivariate Behavioural Research*, *33,* 545-571.

Statistics Canada. (1993). *Northwest Territories' Health Attitudes, Knowledge and Behaviours Study*. Ottawa, ON: Special Surveys Division, Statistics Canada.

Statistics Canada. (2006). *Microdata user's guide: National longitudinal survey of children and youth, Cycle 6*. Ottawa, ON: Special Surveys Division, Statistics Canada.

Statistics Canada/Human Resources Development Canada. (1996). *National longitudinal survey of children and youth users handbook and microdata guide (Cycle 1, Release 2, Microdata document 89M0015 GPE)*. Ottawa, ON: Special Surveys Division, Statistics Canada.

Statistics Canada/Human Resources Development Canada. (1999). *National longitudinal survey of child and youth cycle 3 survey instruments 1998-99: Book 1 - parent and child*. Ottawa, ON: Special Surveys Division, Statistics Canada.

Statistics Canada/Human Resources Development Canada. (2003). *National longitudinal survey of child and youth cycle 4 survey instruments 2000-2001: Book 1 – parent, child & youth*. Ottawa, ON: Special Surveys Division, Statistics Canada.

Statistics Canada/Social Development Canada. (2005). *National longitudinal survey of child and youth cycle 5 survey instruments 2002-2003: Book 1 – parent, child & youth*. Ottawa, ON: Special Surveys Division, Statistics Canada.

Statistics Canada/Human Resources and Social Development Canada. (2006). *National longitudinal survey of child and youth cycle 6 survey instruments 2004-2005: Book 1 – parent, child & youth*. Ottawa, ON: Special Surveys Division, Statistics Canada.

Steiger, J.H. (2000). Point estimation, hypothesis testing, and interval estimation using the RMSEA: Some comments and a reply to Hayduk and Glaser. *Structural Equation Modeling*, *7*, 149-162.

Thomas, E.M. (2007). *Readiness to Learn at School Among Five-year-old Children in Canada*. Children and Youth Research Paper Series. Statistics Canada catalogue no. 89-599-MIE — No. 004. Ottawa, Ontario: Statistics Canada.

Tremblay, R. E., Pihl, R. O., Vitaro, F., & Dobkin, P. L. (1994). Predicting early onset of male antisocial behavior from preschool behavior. Archives of General Psychiatry, 51, 732-739.

World Health Organization. (1997). World Health Organization Survey on Health Behaviours in School Children. Geneva, Switzerland: WHO.

Wothke, W. (2000). Longitudinal and multi-group modeling with missing data. In T.D. Little, K.U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multi-group data: Practical issues, applied approaches, and specific examples* (pp. 1-24). Mahwah, NJ: Erlbaum.

Yuan, K.-H., & Bentler, P.M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. In M.P. Becker (Ed.), *Sociological methodology 2000* (pp. 165-200). Boston: Blackwell Publishers.