Introducing a School Preparedness Index for a Canadian Sample of Preschoolers without Special Needs

Vijaya Krishnan, Ph.D.

Early Child Development Mapping Project (ECMap), Community-University Partnership (CUP),
Faculty of Extension, University of Alberta, Edmonton, Alberta, CANADA

ECMap
Early Child Development
Mapping Project Alberta

Abstract

This paper is an attempt to offer a combined index comprising the five dimensions-physical health and wellbeing, social competence, emotional maturity, language and cognitive development, and anxiety and fearfulness- using a weighting system. The very idea of incorporating the domains of Early Development instrument (EDI) into a composite index was intended to elucidate significant differences of developmental performances in preschoolers across communities. The study included data on 7938 children with no special needs based on kindergarten teachers' responses to 103 questions on a child's behavior in five domains of development, collected in Wave 1 (2009). The component parts of the index were developed from 71 items, using Principal Components Analysis (PCA) with orthogonal rotation. Together, the 71-item version of the five domains accounted for 47.88% of the variance in the data. The composite was constructed using a method of linear aggregation of five components by assigning weights based on the proportion of variance accounted for by the components. The proposed multidimensional index may provide a better picture of child development and stimulate public awareness. However, there are important considerations involved in constructing an index, including whether the methodological and theoretical underpinnings are taken into account, varying perceptions of the societal importance of children's development are addressed, and how to convey the information to both decision makers and the general public, without the loss of any meaningful information.

Keywords Early Development Index, Principal Components Analysis, Composite Index

Introduction

One increasingly popular approach used to understand children's development at pre-school ages involves the use of a rating system known globally as the Early Development Instrument (EDI), developed at the Offord Centre of Child Studies, McMaster University in Canada (Janus & Offord, 2007). It is based on an inventory of 103 questions that a teacher can use to rate a child's behavior in five domains of development: physical health and well-being, emotional maturity, social competence, language and cognitive development, and communication and general knowledge.[1] As currently conceived, the EDI is a multidimensional instrument composed of five quantitative domains, used alone or in combination (as in the vulnerability measure). Two types of measures, interval and categorical, are derived from the EDI: (1) an interval-level measure for each domain, which varies from 0 (low skill/ability) to 10 (high skill/ability), treating the mean of the items contributing to each domain as a domain score; and (2) a categorical measure, the vulnerability score, which is calculated based on a comparison of children's scores with the lowest 10^{th} percentile boundary for each domain. Thus, if a child's score falls below the lowest 10^{th} percentile in one or more domains, a score of 1 (vulnerable) is given, otherwise, a score of 0 is given (not vulnerable).

The five domains vary in terms of the number of items, and in some cases include redundant items, in terms of correlations (Krishnan, 2011). In reality, no tool is able to offer a perfect evaluation of the degree to which a child experiences difficulty or no difficulty at all. To our knowledge, no multivariate study of this kind has been performed to this date addressing the underlying structure of the EDI domains and employing those structures to form a composite index. A short version of the instrument, if reliable and valid, can be more cost-effective and beneficial in large multi-purpose studies or surveys that include other areas of children's health and well-being. The basic aim of this paper was to develop a set of inter-correlated items into a meaningful set of non-overlapping groups based on information obtained from teachers' assessment of children's behaviors. To this end, a Principal Components Analysis (PCA) was performed on the data. As a supplement to the dimensions such as physical health and wellbeing, a composite index was designed to reflect the complexity and multidimensional nature of development; the dimensions were combined to produce a weighted School Preparedness Index (SPI).

[1] UNICEF developed a simple 18-item version of EDI that asks parents to rate their children's behavior in the five developmental domains (Fernald, Kariger, Engle, & Raikes, 2009). CARE employs a simplified version of developmental domains with only three domains–physical, cognitive, and socio-emotional–which involves motor, sensory, language, psychological and emotional aspects (CARE, USAID, Hope for African Children Initiative (2006)).

The Importance of a Weighted School Preparedness Index (SPI)

The idea of creating a single index as a compilation of domains arose from the need to promote and develop policies and programs targeted toward preschool children at the community level. Despite their many deficiencies, discussed later, composite indices can be useful as a communication tool by allowing quick comparisons of community performance in terms of school preparedness. A single index may offer decision-makers a condensed and simple measure of early childhood development without significant loss of information. An index with its weighted sub-components (and individual items), if based on solid theoretical and methodological underpinnings, will prove useful in making an objective comparison of outcomes across various settings. Additionally, it can be linked to other well-known composite indicators, such as Gross Domestic Product (GDP), Gini Index, or Human Development Index (HDI), especially at a national or international level.

What is our rationale behind a composite index of school preparedness? A community, region, or nation needs to know how it is performing in terms of its preschoolers' ability to meet certain standards in their developmental areas. Without this information, it is difficult to rationally plan and/or monitor progress in developmental outcomes, at various levels and sectors. More specifically, by having a composite index, progress can be measured, monitored, and regularly debated and reviewed by policy makers and planners. In addition, this can lead to initiatives to standardize the collection of indicators or variables across states/provinces so their progress can be monitored and tracked against a benchmark. This is what prompted several initiatives in the national and international scenes devising indices, such as GDP, Total Wealth Indicator, Environmental Sustainability Index, to name a few.[2]

Currently, a categorical measure, the vulnerability score, is derived from the EDI based on a comparison of children's scores within the lowest 10th percentile boundary for each domain. Thus, if a child's score falls below the lowest 10th percentile in one or more domains, a score of 1 (vulnerable) is given, otherwise, a score of 0 is given (not vulnerable). The interpretation of this measure is complicated for a couple of reasons: First, there is this assumption that the domains are all equal in importance. Cultural (e.g., language) barriers along with socio-economic disadvantages can produce different developmental trajectories. Second, gender differences in developmental outcomes are not taken into account despite there is growing evidence that gender-specific differentials in developmental outcomes exist (see Buchmann, DiPrete, & McDaniel, 2008). A weighted composite may be better suited to draw differentials in developmental outcomes between groups.

[2] Patterson (2002) noted that composites can have a highly influential impact on decision making and on public policy, providing the example of the labor market policy in the United States; United Sates had no labor market policy until the unemployment rate was codified into the USA Statistical Framework in the 1940s and 1950s.

A composite index and its components represent the two sides of a coin. They, however, are not static; they are the results of the interplay between macro and micro-level factors that influence the child's capacity to perform well in the wider society. Therefore, regardless of whether or not the domains can be tracked for changes, they still can provide incomplete pictures of overall performance. They need to be aggregated into a composite index to judge the overall performance of children at an aggregate level. If every child is given the rights and opportunities to be all he or she can be, either through deliberate policies and programs or by any indirect means, a positive outcome may be achieved, but in varying degrees. Stated another way, development, in general, is an ethical ideal. If all children develop the same way, development in one domain might correspond to development in other domains at more or less the same pace. In practice, however, achieving the full potential of social competence, for example, can imply a certain level of underdevelopment in another area of development, mainly due to the diversity of the population and the conditions that foster (non)development. It is this very diversity of children's development and also the difficulty in gaining an overall picture of development that makes it a necessity to develop an index, adjusted for the relative contribution by its constituent parts.

Prior to developing a composite index, however, some thought should be devoted to the various theoretical and methodological challenges in constructing it. This is useful in sending warning signals or precautions against overgeneralizations of results. To this end, a brief description of the conceptual and methodological obstacles to creating a composite index is attempted, regardless of whether or not the index is built upon a sound empirical foundation.

Theoretical and Methodological Issues in the Construction of an Index

An effective strategy for promoting and developing public policies targeted at early years of a child's development is to develop a statistically valid summary index integrating a range of indicators pertaining to health, food and nutrition, child rights/protection, socioeconomic status as well as key aspects of development (e.g., physical, psycho-social). In highly complex, highly dynamic, physical, cultural, social, and economic landscapes, the holistic needs of a child must be considered with interventions at various levels (e.g., individual child, primary family/caregiver, ECD centers within communities, and local authorities/community leaders), accompanied by policies, laws, and action plans at all levels of governments.[3] Yet, such summary indices can be criticized on the ground that the indicators of development, especially

[3] Refer to CARE, USAID, Hope for African Children Initiative (2006) on CARE's 5x5 model outlining the five areas of intervention. Although the focus in the document is mainly on children in Africa and Asia, the CARE model can be useful in promoting the wellbeing of children in any resource constrained setting, especially in the First Nations communities in Canada.

that of children, change quickly over time and space, rendering current indicators useless in the future.

A single index that reduces all indicators to one number can be appealing to users and practitioners alike. Regardless of its appeal, like some of the multifunctional electronic gadgets now in the market, it is impossible to address the varied and interdependent needs of very young children with a single index based on one area of intervention, such as schooling or education. It may not be sufficient to draw sophisticated policy decision making because the concept encompasses a variety of areas requiring policy interventions. It goes without saying, while no index is perfect, if it is poorly constructed or easily misinterpreted, it can send misleading messages and inappropriate policy prescriptions (Readers may refer to: European Commission-JRC (2008); Nardo, Saisana, Saltelli, & Tarantola, 2005a; Nardo, Saisana, Saltelli, Tarantola, Hoffman, & Giovannini, 2005b; Saisana & Tarantola, 2002 for the pros and cons of composite indicators).

Although a study of this kind cannot tackle or even fully discuss the many challenges that go into the construction of a composite index, it is essential to briefly outline some of the methodological issues.[4] First, the strength of a composite is largely dependent upon the quality of underlying variables; the better their analytical strengths, measurability, and theoretical relevance to the construct being measured, the stronger the composite they represent (OECD, 2003). Second, the assignment of weights to the sub-indices or components is a major challenge. Composite indices have been criticized for their deficiencies, in particular the subjective nature of the weighting and aggregation procedures by which the sub-indicators or components are combined. The issue of weights – should some sort of weighting to be applied, if yes, how are the weights determined, and how can the weights account for the very nature of child development as a socially constructed concept – will be given further consideration later in this paper. We note here that their relevance needs to be assessed in terms of the constituent parts that are glued together to build the composite (Nardo et. al., 2005a, 2005b). The reliability of a composite index can be improved by giving the largest weight to the component having the largest overall significance as determined by theory and/or empirical evidence. The use of factor loadings as weights is the norm in sociological research dealing with composite indices. Finally, researchers are often confronted with the uncertainty that creeps into their choice of the aggregation procedures. Aggregation methods should depend upon the measurement unit. A linear aggregation is often cited as a preferred strategy over a geometric aggregation, provided the components are in the same ratio-scales (e.g., Nardo et al., 2005a).

[4] A detailed exercise on the tools as well as challenges to composite indices building has been described elsewhere (Nardo, et. al., 2005a, 2005b).

In addition to methodological challenges involved in developing a composite index, there are theoretical, and in particular population issues, worth mentioning. The child population is not a homogenous group. Gender, economic, familial, and socio-cultural experiences bring diversity in their early childhood years. Individual children's abilities, to a great extent, are explained by their environments. Unquestionably, an index of wellbeing or development must incorporate several interdependent indicators of vulnerability including age, sex, ethnicity, race, poverty, and disability. If performance measures between or among groups are explored, an emphasis for development policies favoring a particular group can be attempted. That is, the dimensions of physical health and wellbeing, social competence, emotional maturity, language and cognition, and communication and general knowledge have significant elements of intra-child configurations, which need to be addressed in order to make the dimensions and indices "effective", in a real sense. Creating a context-specific composite will improve our understanding of the processes underlying children's development, and such an exercise is beyond the scope of a single study. The scope of this paper extends from the identification of underlying constructs in EDI measurements to derivation of a weighted index, which combines all relevant aspects of development in a way that is still meaningful to conceptualize and communicate. This may prove to be useful for comparisons of performance at a community level and can be a starting point for program and policy initiatives.

Composite indices, in general, should be constructed with a scientific basis after a careful analysis of the uncertainties included in their development so that they can produce meaningful and robust policy messages at a local or national level. With this exercise, our hope is that the composite and the processes that lead to a grand index construction would provide guidance to researchers in analyzing the EDI data more effectively.

Methods

Data

The Early Child Development Mapping Project (ECMap) Alberta, supported by the Ministry of Education, Alberta, Canada, is following a province-wide survey of preschoolers, which began in 2009. The questionnaires assessing children's development included 103 questions, kindergarten teachers being the assessors.

The data set for this study came from the EDI Wave 1 (2009) data, covering the developmental aspects of 9641 kindergarten children in Alberta, Canada. Restrictions to include only those children who were in class more than one month, had no special needs, and had scores missing in no more than one domain brought the sample size to 7938. The study includes five waves of data collection, the last being 2013. Wave 1 data consisted of children who were disproportionately city dwellers (84 percent versus 16 percent). The reader is cautioned about this limitation in generalizing the findings from this study to other jurisdictions.

Statistical Analysis

Factor Analysis (FA) is a widely used statistical procedure in the social sciences. There is a general consensus that the technique is preferable to the Principal Components Analysis (PCA) mainly because FA seeks the least number of factors which can account for the common variance shared by a set of variables. Factors reflect the common variance of the variables, excluding unique (variable-specific) variance. That is, it does not differentiate between unique variance and error variance to reveal the underlying factor structure (e.g., Bentler & Kano, 1990; Costello & Osborne, 2005).[5] In contrast, PCA accounts for the total variance of variables. Components reflect the common variance of variables plus the unique variance (Garson, 2010). Nevertheless, PCA is thought to be ideal in the development of composite indicators (Nardo et al., 2005a; Nicoletti, Scarpetta, & Boylaud, 2000). It is easy to use and allows the imputation of weights according to the importance of sub-components or indicators. The use of a PCA is further justified by the fact that the basic aim is to reduce the complex set of 103 items included in the EDI into a set of fewer uncorrelated components to create a grand index of children's development.

First, we explored how well items group under each domain when they were subjected to PCA. Second, after a satisfactory model had been created, a composite index was developed with the use of the factor scores. The index scores were then used to locate each individual child on each of the EDI domains, and thereby on the composite index. This step was necessary before we assign the scores to communities or report them as area measures. Readers are cautioned, however, that items chosen for one context might not be appropriate for assessing the domains, and consequently the composite in other circumstances, for reasons such as data quality, representation, and sample size.

As a first step, the items were entered into a correlation matrix and a Varimax orthogonal rotation with Kaiser normalization was applied to the solution. This procedure generated 17 components with eigenvalues greater than 1.0. The 17 components accounted for 62.3% of the variance in the dataset. However, 23 items loaded on more than one component (four items even

[5] PCA is not a model based technique and involves no hypothesis or assumed relationships between components. FA, on the other hand, is a model based technique, takes into account the relationships between indicators, latent factors, and error. The technique is believed to yield consistent results mainly because of its recognition of error. FA has the ability to show unique item variance, whereas PCA identifies all variance equally without regard to types of variance (shared, unique, and error).

had loadings on three components), and one item did not load on any of the components.[6] The large number of components and cross-loading items made the task of describing the components extremely difficult, especially because the number of domains originally published by the EDI developers is only five. After examining the Scree plot, a decision was made to keep only five components. When the number of components extracted was limited to five, the variance accounted for was reduced to 44.44% from 62.3%. This analysis yielded 18 items with cross-loadings and eight with no loadings. A test of the 77 items after dropping the 26 items resulted in three items with cross-loadings and one with no loading, producing a variance of 46.96%. A test with 73 items after dropping the four produced a variance of 47.53% and two cross-loading items. Finally, a clean solution emerged with 71 items. With a KMO of 0.96, the variance accounted for by the 71 items was 47.88%, almost 4% more than the variance accounted for by all 103 items.

Results of PCA

Tables 1A to 1E present the final run of the five component loadings involving 71 items, derived from PCA when rotated (Varimax) to a simple structure. The widely accepted domains, developed by the Offord Centre and the 5-component solution were compared for their structures in terms of size and item similarity. The physical health and wellbeing domain with its original 13 items was reduced to 6 items (component #4), explaining 5.81% of the variance. The social competence domain with its original 26 items was reduced to 23 items (component #1), with 10 items in common with Offord's. The remaining 13 items actually belonged to Offord's emotional maturity domain. The 23-item component explained 15.61% of the variance. The emotional maturity domain with its original 30 items was reduced to a 10-item structure (component #3) with eight items common to both PCA and Offord classifications. The component explained 8.75% of the variance. The 26-item language and cognitive development domain came closer to component #2 with 24 matching items, producing 12.38% of explained variance. Surprisingly, the 8-item communication and general knowledge domain had no matching component in PCA. Instead, the sub-domain, labeled as anxious and fearfulness behavior by Offord appeared as component #5 in PCA.

[6] The results of these analyses were reported elsewhere (See, Krishnan, 2011). This work also provides a detailed description of the Offord's five domain structure and that obtained from the Wave 1 data for Alberta.

Table 1: Comparing Offord's Domains and PCA's Components (Varimax)

Table:1A

Offord (13 items)	PCA (6 Items)	
Physical health & Wellbeing	**Component #4**	**Loadings**
Well coordinated (Qa08)	Well coordinated (Qa08)	0.437
Proficient at holding pen (Qa09)	Proficient at holding pen (Qa09)	0.747
Manipulates objects (Qa10)	Manipulates objects (Qa10)	0.810
Climbs stairs (Qa11)	Climbs stairs (Qa11)	0.803
Level of energy (Qa12)	Level of energy (Qa12)	0.687
Overall physical (Qa13)	Overall physical (Qa13)	0.805
Dressed inappropriately (Qa02)		
Too tired (Qa03)		
Late (Qa04)		
Hungry (Qa05)		
Washroom (Qa06)		
Hand preference (Qa07)		
Sucks thumb (Qc58)		

Table: 1B

Offord (26 Items)	PCA (23 Items)	
Social Competence	**Component #1**	**Loadings**
Cooperative (Qc03)	Cooperative (Qc03)	0.58
Follows rules (Qc05)	Follows rules (Qc05)	0.707
Respects property (Qc06)	Respects property (Qc06)	0.723
Self-control (Qc07)	Self-control (Qc07)	0.754
Respect for adults (Qc09)	Respect for adults (Qc09)	0.692
Respect for children (Qc10)	Respect for children (Qc10)	0.729
Accepts responsibility (Qc11)	Accepts responsibility (Qc11)	0.692
Takes care of materials (Qc16)	Takes care of materials (Qc16)	0.598
Follow class routines (Qc24)	Follow class routines (Qc24)	0.577
Adjust to change (Qc25)	Adjust to change (Qc25)	0.470
Overall social/emotional (Qc01)	Gets into fights (Qc37)	0.655
Gets along with peers (Qc02)	Bullies or mean (Qc38)	0.681
Plays with various children (Qc04)	Kicks etc (Qc39)	0.635

Self-confidence (Qc08)	Takes things (Qc40)	0.602
Listens (Qc12)	Laughs at others (Qc41)	0.585
Follows directions (Qc13)	Restless (Qc42)	0.691
Completes work on time (Qc14)	Distractible (Qc43)	0.643
Independence (Qc15)	Fidgets (Qc44)	0.651
Works neatly (Qc17)	Disobedient (Qc45)	0.765
Curious (Qc18)	Impulsive (Qc47)	0.773
Eager new toy (Qc19)	Difficulty awaiting turns (Qc48)	0.740
Eager new game (Qc20)	Can't settle (Qc49)	0.661
Eager new book (Qc21)	Inattentive (Qc 50)	0.601
Independent solve problems (Qc22)		
Follow simple instructions (Qc23)		
Tolerance for mistakes (Qc27)		

Table: 1C

Offord (30 Items)	PCA (10 Items)	
Emotional Maturity	**Component #3**	**Loadings**
Help hurt (Qc28)	Help hurt (Qc28)	0.784
Clean up mess (Qc29)	Clean up mess (Qc29)	0.771
Stop quarrel (Qc30)	Stop quarrel (Qc30)	0.776
Offers help (Qc31)	Offers help (Qc31)	0.793
Comforts upset (Qc32)	Comforts upset (Qc32)	0.855
Spontaneously helps (Qc33)	Spontaneously helps (Qc33)	0.795
Invite bystanders (Qc34)	Invite bystanders (Qc34)	0.784
Helps sick (Qc35)	Helps sick (Qc35)	0.839
Upset when left (Qc36)	Eager new toy (Qc19)	0.330
Gets into fights (Qc37)	Eager new game (Qc20)	0.335
Bullies or mean (Qc38)		
Kicks etc. (Qc39)		
Takes things (Qc40)		
Laughs at others (Qc41)		
Restless (Qc42)		
Distractible (Qc43)		
Fidgets (Qc44)		

Disobedient (Qc45)		
Temper tantrums (Qc46)		
Impulsive (Qc47)		
Difficulty awaiting turns (Qc48)		
Can't settle (Qc49)		
Inattentive (Qc50)		
Seems unhappy (Qc51)		
Fearful (Qc52)		
Worried (Qc53)		
Cries a lot (Qc54)		
Nervous (Qc55)		
Indecisive (Qc56)		
Shy (Qc57)		

Table: 1D

Offord (26 Items)	PCA (24 Items)	
Language & cognition	**Component #2**	**Loadings**
Interested in books (Qb09)	Interested in books (Qb09)	0.369
Interested in reading (Qb10)	Interested in reading (Qb10)	0.55
Identifies letters (Qb11)	Identifies letters (Qb11)	0.673
Sounds to letters (Qb12)	Sounds to letters (Qb12)	0.697
Rhyming awareness (Qb13)	Rhyming awareness (Qb13)	0.645
Group reading (Qb14)	Group reading (Qb14)	0.585
Reads simple words (Qb15)	Reads simple words (Qb15)	0.667
Reads sentences (Qb17)	Reads sentences (Qb17)	0.505
Experiments writing (Qb18)	Experiments writing (Qb18)	0.346
Writing directions (Qb19)	Writing directions (Qb19)	0.501
Writing voluntarily (Qb20)	Writing voluntarily (Qb20)	0.429
Write own name (Qb21)	Write own name (Qb21)	0.426
Write simple words (Qb22)	Write simple words (Qb22)	0.511
Write simple sentences (Qb23)	Write simple sentences (Qb23)	0.410
Remembers things (Qb24)	Remembers things (Qb24)	0.589
Interested in Maths (Qb25)	Interested in Maths (Qb25)	0.582
Interested in number games Qb26)	Interested in number games (Qb26)	0.554
Sorts and classifies (Qb27)	Sorts and classifies (Qb27)	0.545

1 to 1 correspondence (Qb28)	1 to 1 correspondence (Qb28)	0.617
Counts to 20 (Qb29)	Counts to 20 (Qb29)	0.601
Recognizes 1-10 (Qb30)	Recognizes 1-10 (Qb30)	0.662
Compares numbers (Qb31)	Compares numbers (Qb31)	0.653
Recognizes shapes (Qb32)	Recognizes shapes (Qb32)	0.525
Time concepts (Qb33)	Time concepts (Qb33)	0.513
Handles a book (Qb08)		
Reads complex words (Qb16)		

Table: 1E

Offord (8 Items)	PCA (8 Items)	
Communication & GK	**Component #5 (Anxiety & Fearfulness)**	**Loadings**
Effective use English (Qb01)	Upset when left (Qc36)	0.490
Listens-English (Qb02)	Seems unhappy (Qc51)	0.648
Tells a story (Qb03)	Fearful (Qc52)	0.799
Imaginative play (Qb04)	Worried (Qc53)	0.801
Communicative needs (Qb05)	Cries a lot (Qc54)	0.574
Understands (Qb06)	Nervous (Qc55)	0.650
Articulates clearly (Qb07)	Indecisive (Qc56)	0.507
interested in number games (Qc26)	Shy (Qc57)	0.517

In terms of internal consistency of items, the Cronbach's alpha was examined for each component. The physical health and wellbeing domain, comprising 6 items yielded a Cronbach's alpha of 0.88. The social competence domain, comprising 23 items yielded a Cronbach's alpha of 0.95. The emotional maturity domain, comprising 10 items yielded a Cronbach's alpha of 0.93. The language and cognitive domain, comprising 24 items yielded a Cronbach's alpha of 0.91. Finally, the anxiety and fearfulness domain, comprising 8 items yielded a Cronbach's alpha of 0.80. The five components were used in the construction of the composite index.

Construction of the Index of School Preparedness (SPI)

The first criterion established for the SPI was to produce a clean factor structure using the EDI items. The EDI contains 103 items, 71 of which constituted the five components using PCA. Potential components for the composite index came from the 71 items and their factor scores. As it may be clear by now, a grand index may be easier to use by decision makers, but its construction presents a controversial and long-standing methodological problem, leaving a researcher unsure what produces a stable and reliable index. In the absence of a widely preferred method, several judgment calls must be made when constructing composite indices, especially in terms of aggregation methods and assignment of weights. These two issues are discussed further in order to arrive at an objective approach. Also discussed are the different standardization procedures that could help us to interpret the composite better.

Aggregation Procedures

There are debates over the most appropriate method of aggregating components/factors in the construction of a composite. In the context of a discussion on environmental indices, Ott (1978) outlined the use of one of the following approaches (See also, Jollands, Lermit, & Patterson, 2004):

- Summation of sub-indices (linear)
- Multiplication of some or all of the sub-indices (geometric)
- Minimum or maximum of the sub-indices

As earlier noted, whereas a linear aggregation is considered better suited if all the indicators have the same measurement unit, a geometric aggregation is considered better if they are expressed in scales that are dissimilar (Nardo et al., 2005a, 2005b). However, when factor scores from PCA are utilized in the computation of a composite, items that yield negative factor loadings need to be subtracted rather than added in the computation because of the items' negative relationships to the underlying factor (DiStefano, Zhu, & Mindrila, 2009).[7] By retaining the item scales, a linear aggregation makes it easier to interpret. In addition, with a linear aggregation, a researcher can make comparisons across factors even when the number of items per factor differs (DiStefano, et al., 2009).

Unlike linear aggregation, in practice, geometric aggregation requires a community with a low score on one component to have a much higher score on others to improve the overall situation. Thus, in benchmarking exercises, as Nardo et al. (2005a) noted, in general, those areas with low scores prefer linear rather than a geometric aggregation. The authors argued that "the absence of an objective way of determining weights and aggregation methods does not necessarily lead to rejection of the validity of composite indicators, as long as the entire process is transparent" (p.23). Jollands et al. (2004) shared the same argument by noting that an aggregation approach is successful if all assumptions are clearly reported, the methodology is made clear, and the index can be disaggregated into individual components without loss of any information.

[7] Another important issue to be considered relates to the nature of sub-indices. They can be either increasing or decreasing scales (Jollands et. al., 2004). In the first instance, higher values mean worse, and in the second instance, higher means better.

Weighting Procedures for Items and Components

Establishing weights is a major challenge when developing a composite as they have significant impact on the composite and the rankings based on the index. The basis for assigning weights can come from expert opinions to more sophisticated statistical methods. A general tendency among researchers is to use an equal weighting approach, where all indicators are given the same weight. This, however, can pose problems. For example, if individual items are given equal weights and a composite is formed from the components that are derived from those items, the composite can be biased because a component with the most number of items is likely to receive a higher weight in a mathematical sense. In the equal weighting situation, there is also the possibility that when two items with almost equal strengths (as measured by correlations) appear in a unique component (as is usually the case), their contribution to the underlying component/factor can be inflated by doubling their contribution, if highly correlated items are assigned comparatively lesser weights, initially.

As Nardo et al. (2005a) suggested, and as is evident from our component structures, it is desirable to minimize the number of indicators in order to achieve transparency and parsimony, and above all to reduce the cost involved in data gathering. Whatever method one chooses, theoretical, public opinions, or empirical, the process involves some value judgment on the part of the researcher. It follows that the conceptual obstacles in creating composites, especially in terms of the assignment of weights to the components and the items within, is a major cause for concern to those interested in using a single number to measure performance.

Different weighting schemes and standardization methods could yield different results, and an analysis of the index for its sensitivity to weights may be highly recommended. It is easier to give equal weights to the underlying items that relate to a given component. As far as components are concerned, although some components say, for example, *communication and general knowledge*, may be easily amenable from a policy perspective, it is difficult to establish a hierarchy among the five components in terms of their relevance from a policy perspective. From our analyses, we observed that the first three components emerged as stronger than the remaining two. However, the reliability of the components does not differ much to justify giving higher weights to these three. Further, we wanted to avoid under- and over-estimation of a component based on the number of items that is included in the component.[8] In addition there is a general agreement that weights should reflect the contribution of individual items, and PCA's built-in mechanism to account for the highest variation in the data with the smallest possible

[8] Linear aggregation can bias the composite, as it cannot reflect the information of its components. It is implied that the combined impact is more than the impacts of each individual part.

number of items and components offers some kind of objectivity compared to other approaches.[9] Specifically, PCA offers a valuable tool in imputing weights before aggregating items or components.

Before discussing the weighting strategies that we have adopted in the present study, it is important to caution readers that the composite is based strictly on the statistical dimensions of the data rather than on theoretical grounds. Note that with PCA, each component presents a set of items having the highest association with it. The linear combination that explains the maximum amount of variance is the first principal component, another linear combination that explains the maximum amount of the remaining variability makes the second component, another linear combination makes the third, and so on, with each component being independent of the previous ones. The correlations between the items determine the weights that can be estimated. The matrix of the factor loadings or the proportion of the total variance, which is explained by the components (square of factor loadings represent the variance), can be used as weights (Nardo et. al., 2005a; Nicoletti et al., 2000).

A two-stage approach to weighting was used for our purpose. First, we assigned equal weights to the items within a given component to derive the five components. In the second step, the analytical solutions obtained after the Varimax rotation of the five components from PCA were aggregated by weighting each component using the proportion of the explained variance in the data set (Table 2). Specifically, the weights were calculated as:

$$Weight = \frac{Explained\ variance\ for\ the\ component}{Total\ (Cumulative)\ variance\ for\ all\ components\ after\ rotation}$$

[9] Readers are referred to DiStefano et al. (2009) and Nardo et al. (2005a, 2005b) for a detailed discussion on these issues.

Table 2: Explained Variance and Proportion of Explained Variance (Weights) based on PCA Varimax Rotation		
Component	**Explained Variance**	**Proportion of Explained Variance**
1	15.606**%**	0.326
2	12.384**%**	0.259
3	8.745**%**	0.183
4	5.813**%**	0.121
5	5.336**%**	0.111
Total	**47.883%**	**1.00**

Standardization Procedures

After the index was constructed, it was standardized or normalized to make it comparable. Three standardization procedures were applied in the construction of the composite, based on: the standard deviation from the mean score, the distance from the group leader's score, and the distance from the best and worst performers' scores (OECD, 2003)[10]:

(1) *Standard deviation from the mean:* As a widely used standardization tool, the method transforms the distribution to become a standard normal distribution with mean 0 and a standard deviation 1. The positive scores indicate a given score as above average performance and the negative scores indicate a given score as below average performance. This involves computations based on the formula:

$$Standard\ deviation\ from\ the\ mean = \frac{Actual\ score - Mean\ score}{Standard\ deviation}$$

(2) *Distance from the group leader*: The method involves assigning 50 to the top score and then ranking others as percentage points from it. It takes the computational formula:

$$Distance\ from\ the\ group\ leader = \frac{Actual\ score}{Maximum\ score} x\ 50$$

(3) *Distance from the best and worst performers* (the *minimum-maximum* method): As another commonly employed method, it gives the index a range between 0 (laggard with minimum score) and 50 (leader with maximum score). By doing so, for a given score, the index expresses the distance from the overall best performing score and the worst one:

[10] The interpretation here does not imply that the EDI is an individual measure. In order for us to make aggregation to any level, we have to start at the individual level, and that is why the reference is made with an individual focus.

$$The\ minimum - maximum\ method = \frac{Actual\ score\ - Minimum\ score}{Maximum\ score\ - Minimum\ score} x50$$

Applying the first method, the index ranged from -5.60 to 1.35. the second, it ranged from -12.02 to 50.04, and the third it ranged from -.03 to 50.03, with a standard deviation of 7.21, reflecting a wide disparity of values on the composite. All three methods produced perfect correlations suggesting that the composite is not influenced by the standardization procedures, and we are better off using any one of the three methods listed above. We used the *minimum-maximum method* for our purpose.

In Table 3 the distribution of scores by quintiles are presented, ranging from the highly disadvantaged (1st quintile) to the least disadvantaged (5th quintile). If the index is uniformly distributed, the difference in mean index score between adjacent quintiles should be even. However, the difference in means between the first and second groups is higher than any other adjoining quintile.

Levene's test for homogeneity of variances was used to test whether the variance in scores is the same for each of the five groups. Levene's test for homogeneity of variance assumes that the variances of the populations from which different samples are drawn are equal. It tests the null hypothesis that the population variances are equal. If the resulting p-value of Levene's test is less than the critical value, the differences in sample variances are unlikely to have occurred by chance. The results of our analysis showed a significance level of 0.000, a value that is small enough to reject the hypothesis (the probability should be less than 0.05 to reject the null) (Table 3). Thus, the five groups demonstrated considerable socioeconomic variability.

Table 3: Descriptive Statistics of SPI, EDI Alberta, 2009 (N=7938)

Quintile	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower bound	Upper bound
1	1587	28.40	6.14	.15	28.10	28.71
2	1588	38.57	1.61	.04	38.49	38.65
3	1588	42.76	0.86	.02	42.72	42.80
4	1588	45.03	0.51	.01	45.01	45.06
5	1587	46.93	0.77	.02	46.89	46.97
Total	**7938**	**40.34**	**7.19**	**0.08**	**40.18**	**40.50**
Levene's statistic = 1645.46; df1 = 4; df2 = 7933; Sig = 0.000						

Conclusion

In terms of the structure of the EDI domains, the present study showed meaningful but different patterns of domains. Although the patterns are less complex compared to the existing and commonly adopted ones (mainly due to fewer number of items), it cannot be easily summarized because of differing approaches. We eliminated items if they presented cross-loadings or no loadings, and this is in sharp contrast to Offord's way of classifying domains and subdomains. The analysis pointed to the shortcomings of measurement and design in the EDI domains. The PCA procedure provided a valid means of statistically reducing a large number of items to a smaller set of meaningful component items. Reductions in the number of items not only serve to increase the subject-to-item ratio, but also allow researchers to build models for smaller areas and subgroups of populations. It has an additional benefit of reducing the time, cost, and energy involved in gathering data on young children.

A particularly relevant aspect of this study is that the patterns of the EDI domains differ from that of Offord's, to a great extent, for the social and emotional domains. For example, whereas the social competence domain emerged with almost the same number of items, the items themselves differed. That is, the assessment of social and emotional domains may be particularly challenging from the point of view of their stability across populations. It may be that the instrument was developed primarily with a focus on behavioral indicators of early child development that were based on theory and/or expert opinions, and in the process, the inter-correlations and the redundancy of certain items were overlooked.

Caution should be taken, however, when interpreting the components comprising social and emotional dimensions. Though we eliminated items that had cross-loadings or no loadings, the items that were removed may represent important aspects of underlying socio-emotional dimensions. Further research will obviously be required in order to establish the usefulness of the removed items. Further, we do not rule out the possibility of inter-correlations among domains. For example, one could expect the two domains of social competence and emotional maturity to be strongly related, with no clear break between the two, especially when all the 103 items are used. Large cross-sectional data sets, disaggregated by such background factors as demographic and cultural and whose main goal is to identify clear factor structures using transparent and clear methodologies, will ultimately be necessary to shed additional light on major domains in terms of their patterns and structures.

We believe the present exercise raises a number of issues and directions for future research. First, we believe that one-third of the items in the EDI may prove theoretically useful in understanding early child development but not empirically useful. Second, it is important to move further in conceptualizing the domain of communication and general knowledge; the domain is not adequately addressed in its present form. Third, the conceptualization of anxiety and fearfulness needs rethinking; the importance of this as a domain should not be downplayed. Finally, while the EDI has been used in many countries, little or no reference is made about a single index that might capture a proxy for school preparedness. A single index can be valuable in setting policy priorities and benchmarking or monitoring performance at a population level. Such an index, whatever the aspect of development it reflects, may facilitate communication with community leaders and the general public, allowing them to track the direction of progress across space and time, along with other macro-level determinants (e.g., socioeconomic status) of early child development. As a proxy, a composite can easily be developed using weighting procedures and future research should focus on its utility in decision making. Future work should also include the validation of the index using independent samples. Research should explicate how the composite varies across sexes, children of diverse backgrounds, diverse communities, and how its different components function conjointly to impact school preparedness.

The composite index, being the sum of its parts, can be deconstructed such that the contribution of individual component parts can be identified and analyzed if performance needs to be looked at in terms of domains or items. Composite indices are generally risky without sound theoretical, methodological and empirical bases. Without a theoretical underpinning, even when they are constructed upon a sound methodological and/or empirical basis, composites are half-cooked. Since the composite is a weighted aggregate measure, it may conceal disparities either among components or within a given component. For example, if the score of physical domain is very low, it then becomes necessary to identify those communities that are most affected by lack of programs and services for children in terms of their physical growth. However, wrong policy decisions, often simplistic ones, can be derived from the composite without paying much attention to the components. A logical concern is that it has the potential to mask problems associated with the individual parts if they are ignored. With all its limitations, a single index may provide a ‘snap shot’ that may draw the attention of policymakers, program planners, and school authorities to set their targets. The simple nature of an index may attract public interest, but realizing its conceptual limits, it should be always supplemented with other relevant information; the index should be interpreted alongside the individual components from which it is constructed. As the old *adage* goes, one should not lose sight of the forest for the trees. When evaluating an index, sometimes we cannot help but get caught up in concentrating on one or two of its parts and eventually forget that we have a grand index that has domains and items engrossed in it.

Acknowledgements

The author is indebted to Dr. Susan Lynch (Director, ECMap) for her support, encouragement and helpful comments throughout the course of this study. Research Analyst, Dr. Huaitang Wang's assistance with the preparation of tables is highly appreciated. The author is grateful to Kelly Wiens (former Acting Director, ECMap), Olenka Melnyk (Communications Coordinator, ECMap), and Oksana Babenko (Research Assistant, ECMap) for their editorial comments on an earlier version of the manuscript.

References

Bentler, P. M. & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research, 25(1),* 67-74.

Buchman, C., DiPrete, T. A., McDaniel, A. (2008). Gender inequalities in education. *Annual Review of Sociology, 34,* 319-337.

CARE, USAID, Hope for African Children Initiative (2006). Promising Practices: Promoting Early Childhood Development for OVC in Resource Constrained Settings (The 5x5 Model). Retrieved from www.crin.org/docs/promisingpractices.pdf

Costello, A. B. & Osborne, J. W. (2005). Best practices in explanatory factor analysis: Four recommendations for getting the most from your Analysis. *Practical Assessment, Research & Evaluation, 10(7),* 1-8.

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation, 14(20),* 1-7.

European Commission-JRC (2008). Composite indicators: An information server on composite indicators and ranking systems. Italy: Institute for the Protection and Security of the Citizen. Retrieved from http://composite-indicators.jrc.ec.europe.eu/FAQ.htm

Fernald, L. C. H., Kariger, P., Engle, P., & Raikes, A. (2009). Examining Early Child Development in Low Income Countries: A Toolkit for the Assessment of Children in the First Five Years of Life. Washington, DC: The World Bank.

Garson, G. D. (2010). Factor analysis. Retrieved from http://faculty.chass.ncsu.edu/garson/PA765/factor.htm

Janus, M. & Offord, D. (2007). Development and psychometric properties of the Early Development Inventory (EDI): A measure of children’s school readiness. *Canadian Journal of Behavioral Science, 39(1),* 1-22.

Jollands, N., Lermit, J., & Patterson, M. (2004). Aggregate eco-efficiency indices for New Zealand: a principal components analysis. *Journal of Environmental Management 73***,** 293-305.

Krishnan, V. (2011). A comparison of Principal Components Analysis and Factor Analysis for Uncovering the Early Development Instrument (EDI) domains (unpublished manuscript), Early Child Development Mapping Project (ECMap), Faculty of Extension, University of Alberta, Edmonton, Alberta.

Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005a). *Tools for Composite Indicators Building* (EUR 21682 EN). Italy: European Commission-JRC.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2005b). *Handbook on Constructing Composite Indicators: Methodology and User Guide* (OECD Statistics Working Papers, 2005/3). Italy: European Commission-JRC, OECD Publishing.

Nicoletti, G., Scarpetta, S., & Boylaud, O. (2000). *Summary Indicators of Product Market Regulation with an Extension to Employment Protection Legislation* (Economic department working papers NO. 226, ECO/WKP (99)18). Retrieved from http://www.oecd.org/eco/eco

OECD (2003). Quality Framework and Guidelines for OECD Statistical Activities. Retrieved from www.oecd.org/statistics

Ott, W. R. (1978). *Environmental Indices: Theory and Practice*. Ann Arbor: Ann Arbor Science.

Patterson, M. (2002). *Headline Indicators for Tracking Progress to Sustainability in New Zealand*. Wellington, New Zealand: The Ministry of the Environment.

Saisana, M. & Tarantola, S. (2002). *State-of- the-art Report on Current Methodologies and Practices for Composite Indicator Development* (EUR 20408 EN). Italy: European Commission-JRC.