INFORMING EARLY CHILDHOOD POLICY:

AN ANALYSIS OF THE SENSITIVITY OF A SCHOOL READINESS RISK INDEX TO CHANGES IN INDICATOR SELECTION

By

KRISTA S. SCHUMACHER

Bachelor of Science in Journalism
Oklahoma State University
Stillwater, OK
1994

Master of Arts in Liberal Studies
University of Oklahoma
Norman, OK
2007

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY

May, 2015

INFORMING EARLY CHILDHOOD POLICY:

AN ANALYSIS OF THE SENSITIVITY OF A SCHOOL

READINESS RISK INDEX TO CHANGES IN

INDICATOR SELECTION

Dissertation Approved:

Dr. Jam Khojasteh

Dissertation Adviser

Dr. Laura Barnes

Dr. Kenneth Kickham

Dr. Matt Vassar

Outside Committee Member

ACKNOWLEDGEMENTS

"In everyone's life, at some time, our inner fire goes out. It is then burst into flame by an encounter with another human being. We should all be thankful for those people who rekindle the inner spirit." -- Albert Schweitzer

I am thankful for the many people who have rekindled my inner spirit – academically and personally – throughout my life and especially during my time as a graduate student. I am indebted to Dr. Laura Barnes, my dissertation chair, mentor and friend. She encouraged and challenged me to stretch beyond my comfort zone and grow as a scholar and researcher, for which I am eternally grateful.

My advisor, Dr. Jam Khojasteh, and committee members Dr. Matt Vassar and Dr. Kenneth Kickham, gave freely of their time and expertise and provided valuable feedback. I credit Dr. Kickham for piquing my interest in statistics. His mentorship over the years contributed significantly to my development and confidence as a researcher.

Dr. J. David Knottnerus fueled my interest in the nature of the human condition and the influence of human interaction on the course of one's life. It was an honor to work with him during my brief time as a sociology student. Dr. Deena Wheeler provided helpful statistical recommendations and a contagious enthusiasm for research and evaluation.

Thank you to Connie Schlittler and Dr. Naneida Lazarte Alcalá in the Office of Planning, Research and Statistics at the Oklahoma Department of Human Services for their support of my dissertation and the opportunity to work as a researcher in OPRS. I am proud to be part of a team dedicated to improving the quality of life for Oklahomans of all ages.

I am grateful for the many friends who supported me during this process, especially Cat Liles for being perhaps my most vocal cheerleader.

I would not have made it this far in life without the support of my family. My sister and brother-in-law and surrogate parents, Kim and Bill Schlittler, are two of the most generous, kindest people in the world. Their unfailing love, understanding, support and encouragement throughout my life shaped the person I have become. The words "thank you" do not do justice to my gratitude and appreciation for all they have done for me.

My husband, Derek Wilson, provided endless encouragement, insightful feedback, stellar cooking and daily doses of humor, which helped me retain my sanity and to persevere.

Finally, this work is dedicated to the loving memory of Nancy Sue Schumacher, who left this world unexpectedly and far too soon in fall 2014. Mom, I hope you know I made it.

Name: KRISTA S. SCHUMACHER

Date of Degree: MAY, 2015

Title of Study: INFORMING EARLY CHILDHOOD POLICY: AN ANALYSIS OF THE SENSITIVITY OF A SCHOOL READINESS RISK INDEX TO CHANGES IN INDICATOR SELECTION

Major Field: EDUCATIONAL PSYCHOLOGY: RESEARCH AND EVALUATION

The importance of school readiness to both the future of an individual child and society as a whole has given rise to several state-specific indexes designed to measure county-level risk for starting school unprepared to learn. One such index is the Oklahoma School Readiness Risk Index (OK SRRI), comprised of indicators known to be associated with poor school readiness. Among several indicators considered for the index, the final set was determined on the basis of multivariate statistical methods. Selected indicators related to race/ethnicity, family structure and economics, and child maltreatment. No health-related indicators were included.

A limitation of indexes is that there are no agreed-upon best methods or established theoretical framework of measurement for their construction, which makes indexes subject to specification error. Many indexes are developed using reflective measurement models, which assume indicators reflect a unidimensional latent construct. Indexes, however, follow formative measurement models in which indicators define a multidimensional construct. The use of an erroneous measurement model has considerable implications for policy and resource allocation decisions.

This study examined the sensitivity of the OK SRRI to changes to the indicator set. An alternate index was created that reduced the number of racial/ethnic indicators and included those related to health, such as low birth weight. Indicator selection was guided by a theoretical framework based on transactional/ecological and cumulative risk models of child development, as well as assumptions of formative measurement models. Nearly one-third of Oklahoma's counties experienced considerable shifts in rank from the original to the alternate index. Most increases occurred for counties with high rates on at least one health indicator, while many decreases were among counties with high rates on multiple racial/ethnic-related indicators.

This study demonstrated that changes to the indicator set can change the meaning of a construct, which underscores the significance of the indicator selection process. Given the political nature of indexes, it is imperative that those with a stake in the outcomes be included in these processes. As most indexes related to social constructs are intended to inform policy and resource decision-making, this study has important implications for the field of index construction.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

More than 15 years after the National Education Goals Panel (NEGP) set forth the goal of "Ready to Learn" (NEGP, 1999), school readiness is still on the national agenda. In his State of the Union Address (2013), President Obama detailed his Plan for Early Education, which calls for investing in quality early childhood education and child care to ensure all children enter kindergarten ready to learn. This is critical as considerable research shows that children who begin school without a solid foundation for learning are more likely to experience poor academic outcomes throughout school than their school-ready peers (Bradley & Corwyn, 2002; Halle et al., 2009). Quality early education programs and child care have been shown to effectively increase children's readiness for school, especially among children at highest risk for poor school readiness (Burchinal, Peisner-Feinberg, Pianta, & Howes, 2002; Shonkoff & Phillips, 2000; Tucker, Zayco, Herman, & Reinke, 2002). Limited resources at federal and state levels, however, restrict the extent to which quality early childhood programs can be implemented. Identifying areas where children are at greatest risk for starting school unprepared to learn can inform policy and resource allocation decisions to reach those most in need. For this to happen, the construct of school readiness must be defined, and factors that affect readiness must be determined.

Construct of School Readiness Risk

School readiness is increasingly understood as a relational construct emerging from children's experiences in their homes, communities and schools (Meisels, 1999). These experiences impact age-appropriate development of social-emotional, problem-solving, language, motor and general cognitive skills (Masten & Coatsworth, 1998). Because of the interaction of environment and development, the construct of school readiness is not easily delineated, and researchers have wrestled with how best to define it for over two decades (Vernon-Feagans & Blair, 2006).

Despite general agreement that being prepared for academic success at kindergarten entry is "school readiness at its most abstract level," there is no universally accepted definition of school readiness as a construct (Snow, 2006, p. 16). Part of the issue is disagreement on which domains of child development are most critical for success. In various studies of teacher and parental perspectives of school readiness, teachers were found to place more emphasis on social-emotional skills, physical health, and approaches to learning than on counting or recognizing letters or shapes; however, the reverse has been found for parents, particularly among those of lower educational attainment and income levels (Heaviside & Farris, 1993; Piotrkowski, Botsko, & Matthews, 2004; West, Germino-Hausken, & Collins, 1995).

Even though an exact definition of school readiness is debatable, it cannot be contested that children who are underdeveloped in some way, whether socially, behaviorally, emotionally or cognitively, are likely to be less successful in school than children who start with a solid foundation in all of these domains. Comprehending how children develop in the context of their surroundings is perhaps more important than knowing which developmental domain is most critical for school readiness. A transactional/ecological framework is often used to understand how biology and environment interact to affect a child's developmental trajectory (Bronfenbrenner & Morris, 1998; Sameroff & Fiese, 2000). Adverse situations, such as abuse and neglect, and limiting circumstances, such as poverty, are factors in a child's risk for being developmentally unprepared for school. The cumulative risk model (Evans, 2004) posits that the combined effect of a number of risk factors is a more important indicator of poor developmental outcomes than any single factor.

For this study, school readiness is defined in the context of risk, meaning that young children who experience factors that impede development are less likely to start school prepared to learn than those without these experiences. Therefore, identifying indicators of adverse sociodemographic and economic conditions is essential to operationalizing the construct of school readiness from a risk perspective.

Indicators of School Readiness Risk

Although there is considerable research that provides strong support for the connection between family and social environments and child development (Fantuzzo, LeBoeuf, & Rouse, 2014; Mashburn & Pianta, 2006; Rimm-Kaufman & Pianta, 2000; Rouse & Fantuzzo, 2009; Salzinger, Feldman, Stockhammer, & Hood, 2002; Vernon-Feagans, Odom, Pancsofar, & Kainz, 2008), these studies have been conducted at the individual level, making them ineffective for use in a policy context. Decisions related to policy and resource allocation are best informed by aggregate data that serve as spatial indicators of performance related to a particular issue (Early Childhood Data Collaborative [ECDC], 2014; Nardo, Saisana, Saltelli, & Tarantola, 2005). While aggregating data from individual-level records that can be matched across indicators is the ideal situation, Pennsylvania is currently the only state to have an integrated system of individual-level administrative data from municipal services such as human services and education (ECDC, 2014). This limitation requires the use of population-level data to identify areas where children may be most at risk for starting school developmentally behind.

The National School Readiness Indicators Initiative (NSRII), a consortium of 17 states, set forth 23 core population-level indicators organized by domains that reflect the transactional/ecological framework of school readiness (Rhode Island KIDS COUNT, 2005). Indicators are measured at a spatial level, such as county or state, and include various sociodemographic, health and child development factors as well as access to early child care and education. Examples are enrollment in early child care and education programs, fourth-grade reading proficiency, and rates of teenage pregnancies, child poverty, and infants with low birth weight. Although the NSRII refers to all indicators as readiness indicators, it is important to note the distinction between indicators of readiness (e.g., reading proficiency) and predictors of readiness (e.g., program enrollment, poverty). These predictors are better thought of as indicators of risk, where risk is a quantifiable construct derived from some combination of multiple predictors. These risk indicators can be further divided into factors thought to increase risk (e.g., poverty) and those that protect against it (e.g., early childhood education). This study focuses on factors demonstrated in the literature as increasing risk for being unprepared for school.

Each NSRII state has selected a sub-set of indicators based on state needs and priorities and regularly monitors these indicators in an effort to influence public policy related to school readiness. A few states, such as Louisiana, Pennsylvania, Illinois and Oklahoma, as well as Washington, DC, have extended indicator monitoring by applying the cumulative risk model to create indexes that measure risk for poor school readiness and rank counties according to index scores (Lazarte Alcalá, Salehezadeh, & Schumacher, 2013; Lazarte Alcalá & Schumacher, 2014; Louisiana State University/Tulane Early Childhood Policy and Data Center [LSU/Tulane], 2012; Moodie & Rothenberg, 2011; Pennsylvania Office of Child Development and Early Learning [PA OCDEL], 2012; Thomas et al., 2012).

Indexes consist of different but related indicators that, when combined in some way, produce a summary score used to compare geographic areas on the state of some construct (Nardo et al., 2008; Saisana, 2008; Simpson, 2006; Tate, 2012). Indexes are particularly useful in a policy context as they summarize several pieces of information in a way that makes the resulting message accessible and meaningful to the lay person (Davidson & Lambert, 2001). For example, the Oklahoma Department of Human Services (OKDHS) initially developed the Oklahoma School Readiness Risk Index (OK SRRI) to aid in preparing a proposal for a Race to the Top – Early Challenge Grant offered by the U.S. Department of Education and U.S. Department of Health and Human

Services and continues to update the index for use by policy makers and the general public.

The Problem

Although indexes fill important gaps in monitoring complex social issues, one of the primary challenges for indexes is that there are no established rules for their construction or agreement on a best method (Jones & Andrey, 2007; Saisana & Saltelli, 2008; Tate, 2012; Wong, 2006). Outcomes of measuring the same construct may vary considerably depending on the assumptions that underlie design choices, such as indicator selection and methods of normalization, weighting and aggregation, which makes index construction “an inescapably subjective process” (Barnett, Lambert, & Fry, 2008, p. 106). Because “there is no ‘correct’ method” for developing an index, there will always be opportunities for interested parties to alter an index to convey a particular message (Simpson, 2006, p. 5). As such, it comes as no surprise that the accuracy and significance of policy-related conclusions are often questioned due to concerns about the robustness of an index with respect to methodological decisions and subsequently the reliability of outcome rankings or scores (Nardo et al., 2005; Saisana & Saltelli, 2008). Decisions related to selecting indicators and weights are particularly vulnerable to political challenges (Esty, Levy, Srebotnjak, & de Sherbinin, 2005; Nardo et al., 2005).

Because indexes are intended for use by mostly non-statisticians, transparency is of utmost importance (Nardo et al., 2005; Saisana, 2008; Sharpe & Andrews, 2012). Indexes constructed in a way that is incomprehensible to the public may only invite suspicion and doubt regarding the outcomes; thus, a delicate balance must be maintained between rigor and simplicity. As such, the effects of various design choices on index outcomes should be tested (Nardo et al., 2005). Generally referred to as sensitivity testing, an index is evaluated for the extent to which it is sensitive to methodological changes at one or more stages of development. An index is said to be robust if rankings or scores remain relatively stable across various methods of construction. This is an important step that contributes to the credibility of the index, and index developers are encouraged to include it in publications of index results (Nardo et al., 2005). However, none of the school readiness risk indexes noted above appear to have been subjected to any type of sensitivity testing.

Issues in Indicator Selection

While choices made at all design stages should be evaluated, perhaps the most important is the selection of indicators. Nardo et al. (2008) stress that indicators, which should be selected on the basis of a theoretical framework, largely determine the quality of an index. Therefore, indicators should be examined for the extent to which they represent both the overall intended construct and the dimensions theorized to comprise this construct.

Abstract constructs, such as risk for poor school readiness, are inherently unobservable on their own; as such, the use of observed variables (e.g., indicators or items) thought to be associated in some way with a particular construct is required. The nature of the relationship between observed indicators and constructs are specified using measurement models. Two predominant measurement models, reflective and formative, make distinctly different assumptions about these relationships. Reflective models are the foundation for scales, which are used when the construct to be measured is unidimensional and items comprising the scale are considered to reflect, or be the effects of, the construct. For example, a test anxiety scale might consist of several items related to concentration and worrying prior to, during and after an exam (Larson, El Ramahi, Conn, Estes, & Ghibellini, 2010). Because these items are believed to be manifestations of the same construct, they are expected to be highly intercorrelated and therefore are interchangeable. Using one indicator over another should not affect the meaning of the results. This is because causal pathways emit from the latent variable (e.g., test anxiety) to the indicators.

Conversely, in formative models, preferable for index construction, a composite variable is formed that is a direct linear combination of the indicators (Bollen & Bauldry, 2011; Grace & Bollen, 2008). For example, the Human Development Index consists of measures of health, education and income (United Nations Development Programme [UNDP], 2013). Rather than being the effects of human development, the indicators are construed as defining, or causing, the construct. Causal pathways emit from the indicators to the composite variable; indicators are not expected to be highly interrelated as each forms a distinct dimension (e.g., health or income) of the construct. As such, indicators in a formative model are not interchangeable. Therefore, in index construction, “omitting an indicator is omitting a part of the construct” (Bollen & Lennox, 1991, p. 308). This makes the process of selecting indicators the most important step in index development.

The fact that the basic assumption regarding the interrelatedness of indicators in a reflective model does not hold for a formative model has serious implications for assessing the quality of an index. Concepts of dimensionality (e.g., “factorial unity”), validity and reliability all assume the use of a reflective measurement model (Bollen & Bauldry, 2011; Diamantopoulos & Siguaw, 2006, p. 271). Using a reflective model to construct an index is likely to misrepresent the intended construct and lead to policy decisions based on inaccurate conclusions (Bollen & Bauldry, 2011; Hudrliková, 2013). Although multivariate statistical methods are recommended to guide methodological decisions, such as weighting, during index construction, the lack of assumed intercorrelations among indicators means that using such methods as the basis for selecting indicators is likely to compromise the theoretical framework (Nardo et al., 2008; Simpson, 2006). For example, in constructing the Oklahoma index, 18 variables were originally identified for consideration based on a review of the literature with the aim of representing the cumulative risk factor model (Lazarte Alcalá et al., 2013). Final indicator selection from among these 18 variables was driven entirely by multivariate statistical methods. As a result, all indicators associated with health, a domain known to be related to school readiness, were excluded from the final index. This calls into question the extent to which the index represents the full multidimensional nature of school readiness risk.

Purpose of Study

This study is a methodological investigation into the effects of indicator selection on the outcomes of a school readiness risk index. Using the Oklahoma index as a sample index, this study is a form of sensitivity analysis that examines the extent to which modifications to the indicator set change county rankings. The overarching research question asks how the use of different indicators alters the meaning of a risk ranking. Methods follow existing frameworks published in the literature for assessing the impacts of indicator selection on index outcomes (Helton, Johnson, Rollstin, Shiver, & Sprung, 1995; Saisana, 2008; Saisana & Munda, 2008; Schmidtlein, Deutsch, Piegorsch, & Cutter, 2008). The specific questions addressed by this study are listed below.

1. What is the relationship among scores on the overall index, the domains and individual indicators?
2. What is the impact of the indicators and domains on the overall index? In other words, do one or more indicators or domains dominate the index?
3. What is the relative effect of individual indicators and domains on outcome rankings? In other words, how stable are index rankings when individual indicators or domains are removed from the index?
4. To what extent do changes to the indicators and domains affect associations of index rankings with a proxy outcome of school readiness risk?

Significance of Study

This study responded to calls for assessing the potential biases of indexes by examining the relative importance of indicators used in the construction of a school readiness risk index (Fekete, 2009; Saisana, 2008). Ultimately, an index is only the sum of its parts; therefore, the selection of these parts must be transparent and defensible, and the parts themselves should logically sum to the whole. The question of which indicators to use to construct an index and the process of selecting these indicators is critically important if risk is to be assessed in such a way that produces results that are transparent, meaningful, and useful.

This study is significant in that it contributes to the transparency of the OK SRRI and to the field of index construction in general by highlighting issues surrounding the process of indicator selection. While many index developers rely on the assumptions of the reflective measurement model, thus calling into question the meaning of index outcomes as well as validation efforts, this study identified and justified the use of a composite (formative) measurement model for index construction and evaluation. This model informed indicator selection for an alternate index that was used for the study and provided a framework for assessing indicator impact and informed validation efforts.

This study used an established framework for examining the relative importance of individual indicators and the meaning of index outcomes in light of changes to the indicator set (Saisana, 2008; Saisana & Munda, 2008). In this way, this study is a demonstration of how applying such a framework can increase understanding of the extent to which an index is measuring the intended construct. Understanding how an index responds to changes in the set of indicators increases confidence in the interpretation and implementation of the results (Schmidtlein et al., 2008). Because of the political nature of indexes, changes in risk ranks may have important consequences in terms of access to resources. The findings of this study will be important to early childhood program administrators, practitioners, policy makers and other stakeholders with an interest in monitoring school readiness risk factors. It will also be of relevance to other states that use or plan to develop similar risk indexes. Results of this study will be shared with OKDHS, the agency responsible for Oklahoma's risk index.

Limitations and Delimitations

An inherent limitation to indexes is data availability. Data for all indicators must be available at the desired geographic level for the desired timeframe, and should come from databases of government agencies or other credible sources (Fekete, 2009; Simpson, 2006). Ideally, the same timeframe would be used for all variables; however, it may not always be possible to measure population-level data, collected and published by different sources, for the same time periods. To control for potential variation due to using revised data or indicators not previously considered, the data used in this study are the same data used to develop the OK SRRI, for which there is some variation in terms of time periods covered. Further, this study defines risk as a concentration of indicators at the county level, which excludes potentially important variables that may arise if examined at more micro levels, such as community or individual child. It may be, for example, that within larger counties high variability may exist on some or all indicators that is not reflected when data are presented at the county level.

This study is further limited by its focus on the indicator selection stage of index construction to the exclusion of other methodological decisions that must be made at later stages. The study is limited in its generalizability due to its focus on a risk index and data from a single state. Additionally, because there are no hard and fast thresholds from which to gauge the results of sensitivity testing, this process is more descriptive than empirical. In light of problems related to index validation, this study did not aim to validate the index per se but rather to assess whether changes to an indicator set resulted in considerably different associations with a measure proposed to reflect, albeit incompletely, an effect of being unready for school. Unlike grade-level assessments of reading, mathematics and other subjects mandated by the No Child Left Behind Act of 2001, there are no federally mandated kindergarten-entry assessments (ECDC, 2014). For this study, the proportion of entering kindergarteners in each county that scored below proficient on state-mandated school-entry pre-literacy assessments was used to represent the result of being at risk for starting school unready to learn. Given that focusing on one developmental domain to the exclusion of others is reductionist in its disregard for the multidimensional nature of school readiness (Snow, 2006), this is clearly an imperfect measure. Ideally, multiple indicators would be used to reflect being unprepared for school, but literacy is the only developmental domain required to be assessed across Oklahoma at school entry. This approach to validation was meant to be of an exploratory nature and followed Saisana's (2008) method of examining whether enough evidence existed to reject (rather than validate) the index as a measure of school readiness risk.

Definition of Terms

The definitions of key terms used in this study are provided below.

<u>School readiness:</u> "[A] transactional construct from an ecological perspective and is at the intersection of person, process, and context. [It is] not within the child but at the intersection and fit between the child and his/her family and the 'readiness' of the classroom/school to teach that child" (Vernon-Feagans et al., 2008, p. 63).

Risk: An "elevated probability of a negative or undesirable outcome in the future" (Masten & Gewirtz, 2006, p. 24).

Indicator: "[V]alue or group of values that give an indication or direction" (Simpson, 2006, p. 2); "a measure that describes a condition...numbers, percents, fractions, or rates used to paint a picture of a specific outcome or situation" (Rhode Island KIDS COUNT, 2005, para.1).

Dimension: The "highest hierarchical level of analysis and indicates the scope of…individual indicators. For example, a sustainability composite indicator can include economic, social, environmental and institutional dimensions" (Nardo et al., 2008, p. 51).

Index (and composite indicator): "[S]ets of items that are 'causal indicators'…that determine the level of a construct" (DeVellis, 2003, p. 10); "a mathematical combination of individual indicators that represent different dimensions of a concept whose description is the objective of the analysis" (Nardo et al., 2005, p. 7) and measured by a composite score or rank that "is a function of indicators and weights" (Saisana & Saltelli, 2008, p. 251).

Organization of Study

This chapter introduced the problem and described the purpose of the study. Chapter II presents a general overview of indexes, discusses in more detail the study's theoretical framework, and reviews the literature on school readiness risk factors. Chapter III describes the methodology and procedures used to answer the research questions, and Chapter IV presents the findings. This work concludes in Chapter V with a discussion of the outcomes and implications for practice and offers directions for future research.

CHAPTER II

LITERATURE REVIEW

This chapter is organized into three main sections. The first discusses issues in index construction, including the choice of measurement models and why their distinctions are important considerations for index construction. The second describes a theoretical framework of school readiness, and the third reviews the empirical literature on factors identified as core indicators of risk for being unprepared for school.

Indexes: Function and Form

The use of social indicators and aggregated indexes for monitoring social issues began in the U.S. in the mid-1960s when a deeper understanding of emerging social problems was desired (Sheldon & Parke, 1975; Tate, 2012). Indexes, also known as composite indicators, are aggregated measures of individual indicators capable of representing a multi-dimensional social system or construct (Saltelli, Munda, & Nardo, 2006). Because of the complexity of social phenomena, indexes are more effective measures of particular issues than individual indicators alone and are used to convey narratives about social problems that merit intervention (Hudrliková, 2013; Nardo et al., 2005; Saltelli et al., 2006; Tallis, 2005). These narratives emerge from the performance of geographic areas, such as a countries, states or counties, on the index. Each country, for example, is assigned an index score, derived through a mathematical combination of observations, and used for benchmarking and ranking countries according to performance on a particular issue (Nardo et al., 2005; Nardo et al., 2008; Simpson, 2006). By reducing a set of measures to a comprehensive, summary view of a complex phenomenon, indexes can raise awareness of critical issues and foster change (Jacobs, Smith, & Goddard, 2004; Nardo et al., 2005).

Indexes are increasingly being applied to measure complex constructs related to a society's social capital (Saltelli, Nardo, Saisana, & Tarantola, 2005), such as well-being (UNDP, 1990), social vulnerability to natural disasters (Cutter, Boruff, & Shirley, 2003; Davidson & Lambert, 2001; Jones & Andrey, 2007), environmental sustainability (Esty et al., 2005), disaster resilience (Cutter, Burton, & Emrich, 2010), technological achievement (Desai, Fukuda-Parr, Johansson, & Sagasti, 2002), lifelong learning (Saisana, 2008), knowledge as an economic driver (Saisana & Munda, 2008), sustainable growth in the European Union (Hudrliková, 2013; Saltelli et al., 2006) and numerous other issues.

Indicator research in general and indexes in particular have become useful for increasing stakeholder participation in responding to social problems; developing and evaluating policies and interventions; allocating scarce resources; and monitoring change over time (Cutter et al., 2003; Jacob & Willits, 1994; Parris & Kates, 2003; Rossi & Gilmartin, 1980; Simpson, 2006; Wong, 2006). For example, the Social Vulnerability Index (Cutter et al., 2003) examines county-level vulnerability to natural disasters as a means of understanding the social burdens of vulnerability and identifying counties where vulnerability is greatest. Vulnerability ranks have been compared to federal allocations of preparedness resources, which revealed that New Orleans, the most vulnerable city in the U.S., received only 1% of such resources at the time of the analysis (Borden, Schmidtlein, Emrich, Piegorsch, & Cutter, 2007). The United Nation's Human Development Index measures well-being at the country level using indicators of life expectancy, educational attainment and income, and has been used to monitor gains in human progress, particularly in developing countries (UNDP, 2013). The Environmental Sustainability Index (Esty et al., 2005) is used to inform environmental policy decisions by identifying issues related to national environmental protection programs that warrant further attention and highlighting best practices.

Recently in the U.S., there has been an emergence of indexes to measure risk for poor school readiness. Oklahoma, Louisiana, Pennsylvania, Illinois and Washington, DC have created school readiness risk indexes with the aim of drawing attention to counties where children are at greatest risk for starting school already behind. These indexes are designed to inform early childhood education policies and practices and resource allocation decisions related to early education and child care (LSU/Tulane, 2012; Lazarte Alcalá & Schumacher, 2014; Moodie & Rothenberg, 2011; PA OCDEL, 2012; Thomas et al., 2012).

Issues in Index Construction

Index construction is not a straightforward process as there is no correct or best method of development (Saisana & Saltelli, 2008; Tate, 2012; Wong, 2006). Subjective decisions have to be made at various stages of design, and the assumptions that underlie these decisions should be carefully evaluated to avoid questionable analytic rigor (Nardo et al., 2005; Nardo et al., 2008; Saltelli et al., 2006). Moreover, because the target audience for indexes is typically policy makers and the general public, and since numerous individual indicators can be argued to relate to a particular construct, methodological approaches should reflect parity between statistical robustness, simplicity and flexibility (Wong, 2006). A delicate balance between simplifying a particular social phenomenon and providing enough detail to identify differences is a necessary and inherently challenging aspect of index development (Diener & Suh, 1997).

Measurement models. The most critical, but often overlooked and perhaps misunderstood, step of index construction is to specify an appropriate measurement model and justify its use (Coltman, Devinney, Midgley, & Venaik, 2008). Nardo et al. (2008, p. 18) note that a suggested topic for future revisions of the *Handbook on Constructing Composite Indicators* is the relationship between index construction and "the traditional measurement theory developed in psychometrics and in particular the relationship between effect and cause indicators and the statistical tools proposed." As discussed in this section, understanding these relationships is critical to developing an index that produces meaningful and relevant outcomes.

An inherent challenge in the social sciences is measuring abstract constructs that cannot be directly observed. To this end, multi-item measures are usually developed, which require the use of measurement models to specify associations between a construct and a set of observed variables, e.g., items or indicators, believed to be associated with the construct. The use of multiple items in a measure enhances its reliability and validity by increasing the likelihood of adequately identifying a particular construct and reducing random measurement error inherent in individual items (Eisinga, Grotenhuis, & Pelzer, 2013). Theories underlying the association between indicators and construct drive which mathematical techniques can be used to model a construct and which methods of assessing a measure's dimensionality, reliability and validity are most appropriate (Bollen & Lennox, 1991).

Overview of reflective and formative models. Two fundamental measurement models are the reflective and formative models, with the reflective being the most common in the social sciences. This model is the foundation for the development of unidimensional scales, which use multiple items to measure an underlying latent variable, such as personality or attitude. Latent variables serve as "a kind of bridge between observed data and theoretical generalization" (Grace & Bollen, 2008, p. 194). A latent variable such as extroversion could be measured with a scale comprised of items theorized to reflect (be the effect of) extroversion, making these items effect indicators. A scale of extroversion might include items related to sociability, talkativeness and enthusiasm (Gosling, Rentfrow, & Swann Jr, 2003). These items are assumed to be highly related as they are all manifestations of the same latent variable (Bollen &

Lennox, 1991; Coltman et al., 2008; Diamantopoulos & Winklhofer, 2001). By definition, a manifestation is a form that something takes when it appears. Therefore, all items used in the scale are a form of extroversion; they are how extroversion is revealed. Because items are related by the same construct, an increase in extroversion will be accompanied by an increase in all of its associated indicators. This means effect indicators move in relation to each other, making them interchangeable. The inclusion or exclusion of a particular indicator does not change the interpretation of the latent variable (Bollen & Lennox, 1991).

The reflective model, however, is not appropriate for all measures desired in the social sciences. A classic example is socioeconomic status (SES), which could be theorized to be a function of education, occupation, income and neighborhood. An increase in SES does not necessarily imply that all of these indicators will increase; gains in any single indicator will increase SES even if the others remain constant (Bollen & Lennox, 1991). This means the indicators are not required to move in relation to each other to be included in the model. In other words, it is not necessary for indicators to be highly related as each is perceived as a “cause” rather than an effect of SES (Bollen & Lennox, 1991, p. 305; Diamantopoulos & Winklhofer, 2001). Bollen and Lennox stress that the term “cause” carries no particular significance other than indicating a latent variable is determined by the indicators.

It has been argued that the term latent variable does not even apply to formative models as, by definition, a latent variable is something that is hidden or concealed; it exists independently of its indicators (Coltman et al., 2008; Hardin & Marcoulides, 2011). Instead of a latent variable, formative models measure a composite variable, which cannot exist independently of its indicators since it is defined by them (Bollen & Bauldry, 2011; Grace & Bollen, 2008). A simple, albeit contrived, illustration is a house, which at its core consists of a foundation, walls and roof. Therefore, the definition of a house is a structure with all three items; an absence of one is an incomplete house. This is because the associations among these items are not expected to be high. Simply because a builder completes a foundation it cannot be assumed that the walls are up or will be soon. Although a house is clearly an observable variable, this example represents the basic idea behind formative models and composite variables and demonstrates why the difference from reflective models is important. When a composite variable is formed through linear combinations of indicators, which are assigned weights, the composite is "simply a weighted sum of its composite indicators" (Bollen & Bauldry, 2011, p. 265). For example, the composite variable of human development is measured by the Human Development Index, which is a linear combination of health, education and income indicators (UNDP, 2013). Human development, expressed as a ranking, does not exist *in its defined form* independently of its indicators. Provided indicators are not highly correlated, removing one changes the definition of the rankings. Because "the measures produce the constructs, so to speak" (Bagozzi, 1994, p. 332), variation in the indicators leads to variation in the meaning of the composite variable (Diamantopoulos & Winklhofer, 2001; Grace & Bollen, 2008).

Because of this characteristic, the inclusion or exclusion of any indicator in a composite model would result in substantive changes to the meaning of the index

(Coltman et al., 2008). This has been confirmed by several studies that found altering the indicator set used to construct an index had considerable impacts on index outcomes (Chakraborty, Tobin, & Montz, 2005; Houweling, Kunst, & Mackenbach, 2003; Jones & Andrey, 2007; Schmidtlein et al., 2008).When theoretically important indicators are excluded part of the construct itself is excluded (Bollen & Bauldry, 2011; Bollen & Lennox, 1991; Grace & Bollen, 2008). In addition, excluding a necessary indicator biases the coefficients of the remaining indicators, assuming the omitted and included variables are correlated, which in turn affects the weights used to construct the index (Bollen & Bauldry, 2011; Bollen & Lennox, 1991). Selecting the most appropriate indicator set, therefore, depends on using the appropriate model to specify and measure the composite variable (Bollen & Bauldry, 2011; Coltman et al., 2008; Diamantopoulos & Siguaw, 2006; Grace & Bollen, 2008; Roy, Tarafdar, Ragu-Nathan, & Marsillac, 2012).

Implications for reliability, validity and dimensionality. Although it has been argued that indexes follow a formative model, it is not uncommon for reflective models to be used in the process of index construction (Bollen & Bauldry, 2011; Coltman et al., 2008; Diamantopoulos & Winklhofer, 2001; Grace & Bollen, 2008). Studies by business and organizational researchers have found important implications of this mistake (Coltman et al., 2008; Diamantopoulos & Siguaw, 2006). Using an incorrect measurement model “undermines the content validity of the constructs, misrepresents the structural relationships within which these constructs are imbedded, and ultimately lowers the usefulness of … theories” (Coltman et al., 2008, p. 1250). Considerable debate exists, however, about the efficacy of formative versus reflective measurement models and the extent to which the former should be used in light of a "lack of theory underlying formative measurement and a misinterpretation of the early psychometric literature" on validity and reliability testing (Hardin & Marcoulides, 2011, p. 753).

The assumption of the reflective model that individual items are comparable indicators of an underlying construct is at the heart of psychometrics, which "is the science concerned with evaluating the attributes of psychological tests" (DeVellis, 2003; Furr & Bacharach, 2008, p. 8). If indicators are reflections of the same latent variable, then empirical assessment can determine whether the set of indicators used in a particular measuring instrument adequately and consistently represent the intended construct. In other words, tests of reliability and validity, as well as evaluations of the dimensional structure, are based on the reflective model (Furr & Bacharach, 2008).

Reliability. Reliability answers the question of how well an instrument measures what it is intended to measure by assessing the consistency of its results. The concept of reliability is based on classical test theory, which posits that a person's observed score (or response) is equal to the sum of a true (but unknown) score (or response) and measurement error (Furr & Bacharach, 2008). The results of any measure are unreliable to a certain extent due to the presence of error, which can stem from a variety of sources, such as respondent fatigue or poorly constructed measures. Because error is theorized to be associated with observed responses in a reflective model, measurement error can be accounted for and in essence extracted (Coltman et al., 2008). Therefore, reliability can be viewed as a measure's relative lack of error as it represents the "proportion of variance [in a set of items] attributable to the true score of a latent variable" (DeVellis, 2003, p.

27). In other words, if a latent variable is theorized to influence a set of items, reliability can assess how much influence is present (DeVellis, 2003). Various empirical estimates can be used to assess the extent to which scores produced by an instrument (e.g., extroversion scale) are reliable estimates of true scores (or the results that would be found in the absence of error). A common reliability measure is Cronbach's alpha, which assesses internal consistency reliability, or the extent to which a set of items measures a single, unidimensional construct. As item intercorrelations increase, alpha is expected to increase, which means there is less error in the model (Furr & Bacharach, 2008).

In the case of formative models, however, there is no assumption regarding the intercorrelations of indicators; nor is there a theory that proposes an observed score, whether from a single indicator (e.g., education level) or from an index (e.g., SES), is a function of a true score plus error. The concept of error as proposed in the reflective model does not fit the formative model. As such, error terms cannot be estimated if the model is estimated in isolation (meaning there is no effect in the model, only causes). This means measurement reliability cannot be empirically assessed (Bollen & Bauldry, 2011; Diamantopoulos & Siguaw, 2006). As Coltman et al. (2008, p. 9) stress, "One of the key operational issues in the use of formative indicators is that no simple, easy and universally accepted criteria exists for assessing the reliability of formative indicators." Despite this limitation, the *Handbook on Constructing Composite Indicators* (Nardo et al., 2008, p. 26) recommends assessing the internal consistency of an index using Cronbach's alpha, but only when treated as a scale. This reflects the confusion regarding reflective versus formative models abundant in the literature on index construction.

Validity. Validity is the extent to which an instrument measures what it is intended to measure. It is concerned with "whether the [latent] variable is the underlying cause of item covariation" (DeVellis, 2003, p. 49). If so, then items should be similarly related in terms of direction and significance with construct determinants (e.g., causes of extroversion) and consequences (e.g., outcomes of extroversion) (Coltman et al., 2008). Assessments of convergent and discriminant validity, or the extent to which a measure's results are alike or different from other measures of the same or opposite construct, are appropriate in a reflective model, as are tests for criterion validity, or the extent to which a measure can predict an outcome (Bollen & Bauldry, 2011; Coltman et al., 2008). Because there is no expected pattern of intercorrelations among indicators in a composite model, however, their covariance structure cannot be used to judge its validity. As Bollen and Lennox (1991, p. 312) argue, "Without external criteria, a cause induced latent trait is psychologically uninterpretable." Thus, the assumption that indicators have similar relationships with construct determinants and consequences does not hold, and the types of validity assessments discussed above are not appropriate. Despite this limitation, criticisms of the lack of index validation can be found in the literature without a discussion of the limitations of such efforts (Fekete, 2009).

Dimensionality. Another validity assessment concerns construct validity, or the extent to which an instrument measures the intended construct. In other words, construct validity asks whether inferences made from the measure's results are appropriate given the intended construct. The multivariate technique of factor analysis is one approach to assessing construct validity by empirically examining the overall structure, or dimensionality, of a measure. The goal of factor analysis is to represent a set of variables in terms of a smaller set of hypothetical (latent) variable (Tabachnick & Fidell, 2007). The covariance structure of a set of items is used to determine how many latent variables, or factors, underlie the set (DeVellis, 2003). By finding a common factor, extraneous variance (e.g., unique and error variance) among a set of items is removed “because the factor score contains only that part of the indicator that is shared with other indicators” (Coltman et al., 2008, p. 10). Accordingly, for factor analysis to be meaningful for assessing dimensionality, a reflective measurement model is required as there is no assumed pattern of intercorrelations in a formative model. Again, this makes covariance analysis of a formative model irrelevant.

Limitations of multivariate methods for indicator selection. Principal components analysis (PCA), a related but statistically different technique from factor analysis, has been recommended for confirming that an acceptable amount of variance is present in a model and to inform later methodological decisions, such as weighting (Nardo et al., 2008; Saisana, 2008). The purpose of PCA is to see what patterns emerge from all variance, as opposed to only covariance in factor analysis (Tabachnick & Fidell, 2007). PCA reduces a set of individual variables into new, uncorrelated variables, called principal components, that account for all variability in the set using a fewer number of variables. PCA is useful for dealing with multicollinearity in multiple regression in that components can be used as predictors in place of highly correlated variables (Pedhazur, 1997). PCA is only concerned with the statistical structure of a set of items (or indicators) and is not driven by any theoretical notion of how items should fit together.

Like factor analysis, PCA is based on the assumption that indicators are correlated. As discussed earlier, the fact that indicators may not have high intercorrelations, which is actually a desired property of indexes, means statistical factors (or components) may not be appropriate for selecting indicators or defining index domains as straightforward interpretation might not be possible (Houweling et al., 2003; Nardo et al., 2008). The contributions of theoretically important variables that do not move in relation to other variables may be minimized; or variables may have moderate loadings across components rather than a large loading on a single component (Messer et al., 2006; Nardo et al., 2008; Saisana, 2008). Jones and Andrey (2007) noted that high loadings reflect spatial relationships among variables rather than those that most strongly influence a given construct. Saisana (2008) demonstrated how multivariate analysis of indicators comprising the Composite Learning Index (CLI), a measure of lifelong learning in Canadian communities, yielded factors that deviated from the theoretical structure of lifelong learning and lacked meaningful interpretation.

PCA is further limited by its potential sensitivity to outliers, data revisions, the inclusion of new geographic units and small sample sizes (Nardo et al., 2008). Although no definitive rules exist regarding sample size, there are several guidelines. Some argue for absolute sample size thresholds, with Comfrey and Lee (1992) suggesting that 50 is poor while 1,000 is excellent, while others support the use of minimum subject-to-item ratios, such as 10:1 (Nunnaly, 1978), or some combination of ratio and size, such as 5:1 with a minimum sample of 200 (Gorsuch, 1983). Using an average of 10 to 15 variables in an index and all counties in a state in the U.S., many states will have samples too small to support the use of PCA. Moreover, these are not samples of counties but rather populations of counties, and therefore the data is not random but a census. Another limitation is the use of aggregated data that come from different sources and timeframes. There is no way to know whether an individual captured by one indicator is included in another, which makes any correlational conclusions tenuous at best (Thomas et al., 2012). Further, correlations among indicators do not necessarily represent the impact of individual indicators on index rankings (Nardo et al., 2008). Therefore, PCA is useful for confirming that a set of indicators accounts for an acceptable amount of variance and represents diverse aspects of a construct, but the statistical representation may not be as useful or informative as one derived from theory.

As an example, the process of selecting indicators for the OK SRRI relied solely on the results of PCA using 18 population-level indicators identified for consideration and multiple regression analysis. This process is described in more detail in Chapter 3, while the relation of indicators to school readiness risk is discussed later in this chapter. Scores on the four resulting components as well as individual indicators that failed to load on a single component were used as predictors of third grade reading and math proficiency, with percent of students below proficiency in each area used as the dependent variable in separate regression models. The final 10 indicators were selected based on the statistical significance of coefficients for the components and remaining individual indicators. Indicators comprising the three components that were significant as well as a significant individual indicator were included in the index.

The goal in using PCA was to account for as much variance as possible using the smallest number of indicators in order to reduce multicollinearity in the regression models and identify index domains. The use of multiple regression analysis was complicated by the fact that population-level data were used and came from different sources covering different timeframes. Because third-grade reading and math scores were from the 2009-2010 academic year, and most data for the indicators, which were restricted to ages 0 to 5, stemmed from 2007 to 2011, it is highly unlikely that many children were captured in both the indicators and the outcome variables. Moreover, Greenland (2001) stresses that while the assumption of effects following a linear regression model may hold with individual-level data, this is not always the case for aggregated data.

Because PCA is based on item intercorrelations, the three significant components were primarily comprised of indicators with the highest bivariate correlations. For example, the first component was comprised of the Hispanic, English-language learners (ELL) and low maternal education indicators. PCA capitalized on the correlation of Hispanic and ELL (r = .81) and the fact that correlations between these indicators and low maternal education (r = .55 and r = .60, respectively) were slightly higher than correlations between low maternal education and the indicators of poverty and young maternal age (r = .47 and r = .49, respectively), included in the second component. One problem with relying on correlations in studies using aggregated data, also referred to as ecologic studies, is that such studies are subject to ecological fallacy, or the use of aggregated (ecological) data to make inferences about individual-level relationships. A source of bias in ecological studies is that associations of variables seen at an individual level are typically different, in terms of magnitude and sometimes direction, than associations of the same variables found in aggregated data (Hammond, 1973; King, 1997).

The factor matrix published in the first report on the index (Lazarte Alcalá et al., 2013) showed a loading of .68 for low maternal education on the first component (loadings on other components were not listed). Given the correlational structure noted above, it is likely that this indicator had a moderate loading on the second component as well. Moreover, the indicator of no prenatal care, known to negatively impact child development, failed to load with other indicators and therefore was excluded from the component structure. Both of these observations reflect concerns discussed earlier that PCA can minimize the contributions of theoretically important variables, while moderate loadings across components can preclude meaningful interpretations.

Although the Oklahoma index still follows the cumulative risk model, with high rates on multiple indicators resulting in a higher risk ranking, it does not represent the full spectrum of the transactional/ecological model of school readiness. The PCA component that was not a statistically significant predictor in the regression analyses was comprised of health-related indicators. Excluding these indicators, and hence a domain, on the basis of statistical analysis alone means the known relationship between health and child development is ignored and hence the index under-represents the construct of school readiness risk. In developing an index in the field of organizational studies, Diamantopoulos and Siguaw (2006) state explicitly that they did not use any empirical assessments of dimensionality (e.g., PCA) in the process of selecting indicators due to the limitations described above.

Data considerations. For indicators to be included in an index, data should be available across the desired spatial levels and timeframes and from reliable and, preferably, publicly available sources (Nardo et al., 2008; Simpson, 2006). Accuracy of the data and credibility of the sources are critical to ensuring an index is accepted by its intended audience, such as policy makers. As one of the main benefits of indexes is the ability to track trends over time, variables should be selected for which data can be efficiently obtained and are available for most, if not all, geographic areas of interest. Ideally, indexes planned for regular updating should use indicators derived from publicly accessible data sources, such as federal and state databases, to increase reliability, strengthen transparency and reduce production costs (Fekete, 2009; Parris & Kates, 2003; Simpson, 2006). Wong (2006) notes the advantage of using official data sources is their credibility and acceptance by the target audience. This can create barriers to the inclusion of relevant indicators if data are not published or otherwise available on a regular basis. For example, the indicators of ELL and abuse/neglect, factors known to decrease children's school readiness and included in Oklahoma's risk index, are not publicly available and require contacting agencies directly. Therefore, it is sometimes the case that a tradeoff must be made between transparency and full representation of a social phenomenon. Finally, as indexes are intended to inform policy decision-making (Wong, 2006), policy-related variables should reflect variation both across spatial levels and across time (Kickham, 2000).

Once collected, data should be examined for missing observations and outliers and transformed as needed. Although there is no expected pattern of correlations among indicators in a formative model, analyzing correlations is an important step in determining whether the multidimensional nature of a construct is represented. Extremely high correlations among several variables may mean these variables measure the same thing and therefore inadequately represent the multidimensional nature of a particular construct. Further, a high degree of multicollinearity obscures the ability to determine each indicator's influence on the latent variable (Coltman et al., 2008; Diamantopoulos & Siguaw, 2006). It may also introduce double counting, which will skew the weights used for individual domains depending on whether correlated indicators are used in the same domain (Nardo et al., 2008). A threshold for high correlations, however, has not been established, and other indexes have included indicators with correlations as high as .90 (Jones & Andrey, 2007). The key is that there is not a systematic pattern of high correlations among multiple indicators (Saisana, 2008).

Overcoming Issues in Index Construction

Four main points related to indexes emerge from this discussion: 1) the construct to be measured by an index must be explicitly defined; 2) careful selection of indicators based on a strong theoretical perspective of their relation with the construct is imperative; 3) using multivariate techniques to select indicators and identify the domain structure of an index can result in the exclusion of important indicators and misrepresent the intended construct; and 4) commonly used approaches to assessing the psychometric properties (validity and reliability) of a measure are inappropriate for indexes. Bollen and Lennox (1991) sum up the situation related to index construction and evaluation nicely:

> Conventional wisdom on item selection and scale evaluation is thus shown to be qualified by consideration of the specific directional relationship between the indicators and the latent construct. Traditional measures of reliability and the examination of the correlation matrix of indicators are so ingrained that researchers have failed to realize that these are not appropriate under all situations (p. 312).

Validity and reliability, therefore, take on new meaning for indexes. Since reliability cannot be empirically assessed, error in composite models can only be overcome by the design of the study, such as capturing all possible causes of the construct (Bollen & Lennox, 1991; Coltman et al., 2008). Diamantopoulus and Siguaw (2006) recommend examining the extent to which indicators have the same directional relationship with a construct as a means of assessing their appropriateness for an index.

Another approach to assessing the reliability of an index is through sensitivity analysis (Fekete, 2009; Saisana & Saltelli, 2008). Sensitivity analysis aims to assess the extent to which methodological approaches to index construction affect the stability of summary rankings (Li & Wu, 2006; Nardo et al., 2005; Saisana, 2008; Tate, 2012). An index is said to be robust to methodological decisions if its outcomes remain relatively stable across changes to methodology, such as the use of different indicators or weighting schemes (Saisana & Saltelli, 2008; Tate, 2012). Choices surrounding indicator selection and weighting, for example, are sources of uncertainty that can significantly impact outcome scores and rankings. Understanding how such uncertainties proliferate throughout the structure of an index to influence outcomes is a means of assessing the reliability of index rankings (Saisana & Saltelli, 2008). For example, a recommended early step in index construction is to identify the main drivers of performance, which reveals the relative importance of each indicator and whether the index is dominated by particular indicators (Nardo et al., 2008). This is referred to as “decomposition [of an index] into the underlying indicators” (Saisana, 2008, p. 14) and has been accomplished through various methods, including path analysis (Saisana & Munda, 2008), stepwise regression (Helton et al., 1995), and comparing rankings or scores from scenarios that leave out one or more indicators to those from the full index (Chakraborty et al., 2005; Saisana, 2008; Schmidtlein et al., 2008). Even if uncertainty cannot be reduced, analysis of the extent to which an index is sensitive to various sources of uncertainty lends credibility to a model by increasing its transparency is a means of ensuring valid interpretations of index outcomes (Li & Wu, 2006; Saisana, Tarantola, & Saltelli, 2005).

Regarding validity, one perspective of index validation involves determining whether there is enough evidence to reject an index as an appropriate measure (Saisana, 2008). To this end, linking index outcomes with another construct theorized to be either a determinant or outcome of the composite variable and derived through a reflective model is recommended (Bollen & Lennox, 1991; Coltman et al., 2008; Diamantopoulos & Siguaw, 2006; Saltelli et al., 2005). However, determining the direction of causal pathways in a model theorized to be the outcome of an index can be challenging. As Saisana (2008) notes regarding the construction of the Economic and Social Well-Being

Index, constructed using a reflective model in an effort to validate the CLI, the indicators used could very well be theorized to cause a society's well-being rather than be its effect. Therefore, although attempts to validate an index in this way are useful, the results do not carry the same definitive meaning as they might in traditional psychometric testing.

Given these limitations, the definition of valid information used in *The Program Evaluation Standards* seems relevant for index construction. According to the standards, a valid evaluation is one that "serve[s] the intended purposes and support[s] valid interpretations" (Yarbrough, Shulha, Hopson, & Caruthers, 2011). Careful analysis of decisions made regarding indicator selection is a means of understanding the extent to which an index is internally sound. At their foundation, indexes should have a theoretical framework that provides a clear understanding of the multidimensional nature of the phenomenon being measured and the domains represented. This serves as the basis for selecting and combining individual indicators under a "fitness-for-purpose" principle in which the overall index is considered for its relevance to the intended purpose and acceptance by intended users (Nardo et al., 2005; Saltelli et al., 2006, p. 224). As Grace and Bollen (2008, p. 201) stress, "any composite derived from the measures of that construct should be understood to represent the collective effects of its components, regardless of the label placed on that construct." In other words, an index is no more than the sum of its parts, which, depending on the parts used, can convey different meanings.

Theoretical Framework of School Readiness

The transactional/ecological and cumulative risk perspectives discussed in this section serve as the theoretical foundation for the selection of indicators for a school readiness risk index and the testing of varying indicator sets in the proposed study. These approaches naturally complement one another as an overall theoretical framework for understanding how risk factors interact to restrict a child's normal pace of development. A child's environment and experiences can range from providing all resources a child needs for healthy development to posing considerable barriers to development and subsequently school readiness.

Children who are developmentally ready to start formal schooling possess age-appropriate cognitive, social, emotional and physical skills. They are both ready to learn and ready to engage in the school environment (Carlton & Winsler, 1999). Definitions of school readiness that rest on children's cognitive abilities, however, assume the experiences and learning opportunities in the early years are similar for all children (Meisels, 1999). Rather, a child's development is highly dependent on the environment within which a child grows (Glaser, 2000; Kelley, 2003; Rodriguez & Tamis-LeMonda, 2011; Shonkoff & Phillips, 2000; Thompson, Levitt, & Stanwood, 2009). Nurturing early environments contribute to normal brain maturation and appropriate development of cognitive, linguistic, and social/emotional skills (National Institute of Child Health and Human Development, 2002; Peisner-Feinberg et al., 2001; Shonkoff & Garner, 2012). These early skills can significantly impact subsequent achievement, such as performance in later primary grades (Claessens, Duncan, & Engel, 2009), high school completion (Garnier, Stein, & Jacobs, 1997) and college enrollment (Brooks-Gunn, Guo, & Furstenberg Jr., 1993).

Transactional/Ecological Model

With the current interest among researchers, policy makers and community members in school readiness, it is generally accepted that school readiness is a multidimensional construct highly influenced by interrelated factors occurring in the context of home, school and community (Hair, Halle, Terry-Humen, Lavelle, & Calkins, 2006; High, 2008). This perspective reflects the transactional/ecological framework of readiness that views child development as a product of complex interactions of experiences and environment (Felner, Felner, & Silverman, 2000; Huston & Bentley, 2010; Vernon-Feagans et al., 2008). This framework stems from ecological systems theory (Bronfenbrenner & Morris, 1998) and transactional perspectives of development (Sameroff & Fiese, 2000). Together, these perspectives provide a foundation for understanding how circumstances such as poverty and associated adverse situations, such as teenage mothers and poor prenatal care, impact developmental outcomes (Felner & DeVries, 2013).

Ecological systems theory, which serves as the foundation for the federally funded Head Start program for children in poverty, posits that child development stems from the interaction of five expanding structural systems (Bronfenbrenner, 1979; Bronfenbrenner & Morris, 1998). The microsystem is the first system and consists of a child’s immediate surroundings, such as home or school, which most directly impact child development. Following this is the mesosystem, which is the interaction of multiple microsystems in which children are active participants. These interactions are thought to be equally as important as the occurrences within a single microsystem. Extending beyond the mesosystem is the exosystem, which consists of settings in which a child may never be directly involved but that nonetheless influence development. Examples of exosystems include parental employment or availability of child care subsidy benefits. The fourth system is the macrosystem, which consists of the values and beliefs of the culture within which a child lives, such as the value of a two-parent family or the belief that spending most of a child's time in school contributes to success. Finally, the ecological systems conclude with the chronosystem, which encompasses the passage of time and considers the duration of exposure, either directly or indirectly, to experiences within the other four systems.

Ultimately, ecological systems theory is based on the premise of development as "a function of forces emanating from multiple settings and from the relations among these settings" (Bronfenbrenner & Morris, 2007, p. 817). The "proximal processes" that interact between person and environment are key factors in the effects of these ecological systems on development, yet these processes are not the same for everyone and are influenced by the "biopsychological characteristics" of an individual as well as the contexts and time periods within which experiences occur (Bronfenbrenner & Morris, 2007, p. 795). This concept of development as the interaction of biological and psychological traits, environment and time, is related to the proposition that development is transactional in nature. Systems such as the home and school interact over time not only with each other but also with a child's internal processes, such as cognitive and social-emotional skills, to create a particular developmental trajectory (Sameroff & Fiese, 2000). Thus, a transactional/ecological perspective has emerged that considers “relationships between individuals and those [disadvantaged] environments, and the ways in which those environments and their experience may interact with each other, across contexts” to disrupt normal developmental processes (Felner & DeVries, 2013, p. 110).

Cumulative Risk Model

Another perspective, the cumulative risk model, operationalizes the transactional/ecological framework by proposing that it is the accumulation of risk factors that most inhibit a child’s developmental trajectory rather than any single factor (Evans, 2004; Stanton-Chapman, Chapman, Kaiser, & Hancock, 2004). Therefore, the combined effect of a number of factors should be assessed when measuring risk for poor school readiness rather than any single factor, which supports the use of an index to measure risk. Although particular types of risk, such as low maternal education, may inhibit healthy development of cognitive, language and social-emotional skills, type is not as important a consideration in child outcomes as is exposure to numerous risk factors (Burchinal, Roberts, Hooper, & Zeisel, 2000). It is hypothesized that exposure to a single risk factor is less disruptive than multiple risks due to increased stress levels that accompany a multiplicity of negative experiences (Deater-Deckard, Dodge, Bates, & Pettit, 1998).

A review of the literature shows that considerable research has used the cumulative risk model to investigate the extent to which exposure to co-occurring risk factors, particularly during a child’s early years, restricts development and hampers achievement. For example, a longitudinal study of the effects of early and middle childhood exposure to child abuse, inter-parental violence, family disruption, poverty and high parental stress on adolescent behavior showed that child outcomes worsened as children experienced a greater number of disruptive situations (Appleyard, Egeland, van Dulmen, & Alan Sroufe, 2005). Of particular importance to the topic of school readiness is that exposure during the early years explained variation in adolescent behavior even when controlling for exposure during middle childhood. The co-occurrence of risk factors has also been found to negatively impact child health. This is important to understanding how risk exposure impacts school readiness as poor child health is known to be associated with low academic achievement (Eide, Showalter, & Goldhaber, 2010). Research on the effects of poverty, minority race/ethnicity, low parental education levels and single-parent families found that children's health status was significantly worse for children with more than one risk factor (Bauman, Silver, & Stein, 2006). Another study investigating eight social risk factors, including lack of health insurance, family conflict and low maternal mental health, among others, found that while individual factors were modestly related to poorer child health status, children with six or more risk factors were nearly 17 times more likely to be in poor health as those who experienced no risks (Larson, Russ, Crall, & Halfon, 2008). Numerous other studies point to the deleterious effects on academic outcomes of exposure to multiple risk factors above and beyond individual factors (Burchinal, Roberts, Zeisel, Hennon, & Hooper, 2006; Burchinal, Roberts, Zeisel, & Rowley, 2008; Gutman, Sameroff, & Cole, 2003; Sektnan, McClelland, Acock, & Morrison, 2010).

School Readiness Risk Factors

Research suggests that providing children at high risk for poor school readiness with quality interventions at the earliest ages positively affects both cognition and long-term health, which in turn contributes to increases in lifelong achievement and earnings potential (Campbell et al., 2014). Although schools are increasingly held accountable for student achievement, they often lack the resources to adequately address learning deficiencies that stem from detrimental circumstances in the home and community. However, studies of the extent to which school characteristics and resources, such as size and pupil expenditures, explain gaps in achievement have yielded mixed results (Teddlie & Reynolds, 2000). Instead, research has more consistently found that achievement gaps between schools are better explained by certain sociodemographic characteristics of families and communities and the concentration of these characteristics in schools (Fantuzzo et al., 2014).

Efforts such as the NSRII and the State Early Childhood Comprehensive System (ECCS) Initiative have resulted in the identification of numerous indicators related to school readiness and recommendations for regular monitoring of these indicators at the state or local levels. The NSRII is a consortium of 17 states that came together with the goal of informing early childhood policy issues by identifying indicators thought to be most important in estimating school readiness. The NSRII organized the concept of school readiness into 6 domains that reflect the transactional/ecological framework of child development, and 23 population-level indicators nested within these domains were selected as core indicators of school readiness (Rhode Island KIDS COUNT, 2005).

Indicators are measured at a spatial level, such as county or municipality, and address various sociodemographic and health risk factors, as well as access to quality early child care and education and child development outcomes. Each NSRII state is encouraged to assess the state's most pressing needs and identify a sub-set of indicators for regular monitoring.

The NSRII domains are "ready children," "ready families," ready communities," "ready services – health," "ready services – early care and education," and "ready schools" (Rhode Island KIDS COUNT, 2005). "Ready children" measures child development outcomes, including language, social-emotional and cognitive development. "Ready families" considers a child's family environment, particularly maternal characteristics. "Ready communities" reflects an area's socioeconomic status and support services for low-income families with young children. "Ready services – health" includes birth outcomes and other indicators of child health. "Ready services – early care and education" measures access to quality early education and child care and to child care subsidies. Finally, "ready schools" consists of classroom size for young children and fourth-grade reading proficiency.

Another school readiness indicator effort is the ECCS, a national initiative to integrate early childhood services within states and monitor child development risk indicators and outcomes. An assessment of state ECCS reports and reviews of the empirical literature and existing indicator efforts at the state and national levels resulted in the recommendation of 36 indicators of school readiness (Johnson, Davidson, Theberge, & Knitzer, 2008). The assessment was conducted for the National Center for

Children in Poverty in an effort to facilitate cross-state comparisons and begin to monitor early childhood development and school readiness at the national level (Johnson et al., 2008). Recommended indicators are organized into categories that reflect the NSRII domains and include child development and academic outcomes, population-based risk factors, health indicators, social-emotional development, and access to quality early child care and education. Of NSRII's 23 indicators, all except early childhood classroom size are also identified as proposed indicators for monitoring by all ECCS states. The ECCS includes additional population-based risk factors (race/ethnicity and exposure to extreme poverty) not reflected by the NSRII and emphasizes consideration of exposure to multiple risk factors, which reflects the cumulative risk model.

Conceptual Framework for School Readiness Indicators

As NSRII and ECCS indicators address three key components of school readiness – risk factors, processes for ameliorating risk and child development outcomes – the overall concept of school readiness can be broadly conceptualized using a logic model framework. Logic models are frequently used in evaluation efforts and outline "how a program will work under certain conditions to solve identified problems" (Renger & Titcomb, 2002, p. 493). Logic models consist of three main components: "antecedent conditions," or causal factors related to a problem (Renger & Titcomb, 2002, p. 496), responses to these conditions and outcomes. Using indicators recommended by the NSRII and ECCS, a logic model of school readiness would include risk factors as antecedent conditions; quality early education, child care and family support services as responses; and child development and reading proficiency rates as outcomes.

Understanding what is expected to change within a given timeframe is critical when establishing outcomes. As such, two additional population-level outcomes could be addressed in this framework: 1) increases in access to early childhood programs and family support services in areas where risk for poor school readiness is high; and 2) decreases in the prevalence of risk factors over time. This second outcome points to the need for focusing attention on risk indicators that can be changed, and developing responses to these problems through interventions or policies. These responses, such as programs aimed at reducing abuse and neglect by increasing parenting skills, should be included in the second stage of a school readiness logic model.

Table 1 depicts a generic logic model for school readiness. Antecedent conditions are organized using NSRII domains representing risk indicators, which could be useful for organizing indicators to ensure full construct representation in an index. Johnson et al. (2008) propose a similar approach to conceptualizing school readiness using a results-based accountability (RBA) framework that stresses accountability for the well-being of entire populations and client populations in particular. The use of indicators to develop

Table 1

Generic Logic Model Framework for School Readiness Indicators

Antecedent Conditions	Responses	Outcomes
Family risk indicators	Early education programs	Child outcomes, e.g., developmental domains and reading proficiency
Community risk indicators	Child care services	
Health risk indicators	Programs/policies to reduce risk factors	Population outcomes, e.g., reduction in risk factors

strategies for improving quality of life for children and families through programs and policies is a key component of RBA from a public accountability perspective (Friedman, 2005).

Review of School Readiness Risk Indicators

Table 2 lists indicators of school readiness risk recommended for regular monitoring by the NSRII and ECCS and/or included in the school readiness risk indexes created for Oklahoma, Pennsylvania, Louisiana, Illinois and Washington, DC. As there are differences in the organization of indicators into domains and the naming of similar domains across monitoring efforts, the domains used to categorize risk indicators for this study follow those established by the NSRII of “ready families,” “ready communities” and “ready services – health.” Table 2 shows only one indicator – births to teenage mothers – is common to the seven monitoring efforts. Income is also reflected in all efforts, although one index measures it as family income less than 185% of the federal poverty level compared to less than 100% used in the other efforts. In addition to family income below the poverty level, three indicators – low maternal education, child abuse/neglect and low birth weight – are common to at least six efforts. Of the indicators listed, 10 are included in both the NSRII and ECCS recommendations. Table 2 also shows that indicators used for six of the seven monitoring efforts cut across all three NSRII domains represented. The exception is the Oklahoma index, which does not include indicators under the domain of “ready services – health.” The ECCS is the only entity to have an indicator measuring the presence of three or more demographic risk factors, such as poverty and parents who are single, non-English speaking, have less than a high school education or are unemployed. Although much research points to the increased risk among racial/ethnic minorities, only three efforts – the Oklahoma and Illinois indexes and ECCS states – monitor racial/ethnic minority indicators, which are set apart from the NSRII domains in Table 2.

The following section reviews the empirical literature on factors that place children at risk for poor school readiness. Representing the NSRII domains of ready families, communities and health services, these factors are perhaps the most critical as they are the most proximal to a child's development.

Family factors. Several factors related to family structure and environment have been found to affect school readiness, with many factors occurring at a higher rate among children in poverty (NEGP, 1997). Of numerous risk factors, low maternal education may be the most important predictor of poor school readiness. Research using individual-level data found that having a mother with less than a high school diploma was a stronger and more consistent predictor of poor reading and mathematics skills and inconsistent attendance patterns among third-grade students in an urban school than any other risk factor assessed (Fantuzzo et al., 2014; Rouse & Fantuzzo, 2009). Moreover, enrollment rates in early learning programs have been found to decline with maternal education level (Barnett & Yarosz, 2007), which may partially explain the negative effect of under-educated mothers on a child's school performance (West, Denton, & Germino-Hausken, 2000; Zill & West, 2001). Using data from the National Evaluation of Welfare-to-Work Strategies Child Outcomes Study, Magnuson (2003) found that increases in maternal education positively affected school readiness.

Table 2

School Readiness Risk Indicators used in Seven Monitoring Efforts

	NSRII	ECCS	OK	PA	LA	IL	DC
Ready Families							
Births to mothers with < 12th grade education	✓	✓	✓	✓	✓		✓
Births to teenage mothers	✓	✓	✓	✓	✓	✓	✓
Child abuse/neglect	✓	✓	✓	✓		✓	✓
Children in foster care	✓	✓	✓	✓			
Children birth to 5 living with single parent		✓	✓			✓	
Births to single mothers				✓	✓		✓
Ready Communities							
Lead poisoning under age 6	✓	✓				✓	
Children in families with incomes < 100% of federal poverty level (FPL)	✓	✓	✓	✓	✓		✓
Children in extreme poverty (≤ 50% FPL)		✓					
Children in families with incomes < 185% FPL						✓	
Median income as percent of poverty				✓	✓		
Unemployment rate		✓			✓		
Homeless children						✓	
Children in families that receive public assistance						✓	✓
School free and reduced lunch participation				✓			
Children who are English-language learners			✓				
Children of parents who are migrant workers			✓				
Ready Services - Health							
Uninsured children under age 6	✓	✓			✓		
Vaccination rate children 19 to 35 months	✓	✓					
Low birth weight infants	✓	✓		✓	✓	✓	✓
Births to mothers with late or no prenatal care	✓	✓					✓
Preterm births (prior to 32 weeks)				✓			
Infant mortality rate				✓	✓		✓
Tobacco use during pregnancy				✓			
Race/Ethnicity							
Children birth to 6 of non-white race/ethnicity		✓				✓	
Children birth to 4 of Hispanic ethnicity			✓				
Children birth to 4 of American Indian race			✓				
Cumulative Risk							
Children with multiple risk factors (≥ 3 risk factors)		✓					

Note: ✓Items included in risk monitoring efforts.

Several studies have found an association between teenage pregnancies and negative birth outcomes, such as low birth weight and inadequate prenatal care (Chen et al., 2007; Fraser, Brockert, & Ward, 1995). Teenage girls who live in poverty, are a racial/ethnic minority, or immigrated to the U.S. are more likely than others girls to become pregnant, receive little to no prenatal care, and drop out of school (Abrahamse, Morrison, & Waite, 1988; Chandra, Schiavello, Ravi, Weinstein, & Hook, 2002; Fiscella & Kitzman, 2009; McLafferty & Grady, 2004). As of 2010, Oklahoma was among the top five states in terms of births to teen mothers at 50 or more births per 1,000 teenage girls (Hamilton, Martin, & Ventura, 2012; Martin et al., 2012).

Children born to teen mothers are at a particularly high risk of abuse and neglect compared to those of older mothers (Bartlett & Easterbrooks, 2012). Chronic stress from exposure to abuse and neglect presents serious risks for under-development. Children in abusive and neglectful environments are at higher risk for slowed brain development and poor academic performance than those raised in nurturing environments (Knitzer & Lefkowitz, 2005; Shonkoff & Phillips, 2000). Increasingly referred to as "toxic stress," frequent or prolonged triggering of the body's stress response system without the protective benefit of a nurturing adult relationship has been found to permanently change the structure of the brain and its functional capacity (Shonkoff & Garner, 2012). Longitudinal studies have demonstrated that adults who were abused or neglected as children have lower IQ scores, shorter attention spans, poorer memories, and increased risk of dropping out of school than non-abused or neglected children (Erickson & Egeland, 1996; Sapolsky, 1996; Widom, 2000). Severe child abuse and neglect often leads to foster care placement (Stovall-McClough & Dozier, 2004), which is most prominent among the poor (Barth, Wildfire, & Green, 2006) and racial/ethnic minorities (Carter, 2010; Courtney & Skyles, 2003; Needell, Brookhart, & Lee, 2003). Several studies have demonstrated a strong relationship between foster care placement and poor academic outcomes (Fantuzzo & Perlman, 2007; Knitzer & Lefkowitz, 2005; Pears, Fisher, & Bronz, 2007).

Factors associated with being from a single-parent family, such as poverty and decreased parental interaction with children, place a child at high risk of delayed social and academic development (Shonkoff & Phillips, 2000). Of single parents, most are mothers, and research shows that households headed by single mothers are more likely to be impoverished than two-parent households (McLanahan, 2004; Shonkoff & Phillips, 2000). In a meta-analysis of 67 studies from the 1990s, Amato (2001) found that children of divorced parents had significantly lower academic achievement than children of married parents. Being from a single-parent home significantly increases the risk of dropping out of high school, as demonstrated by several national surveys (McLanahan & Sandefur, 1994). For children of single mothers, positive involvement by fathers has been shown to reduce negative outcomes (Coley, 1998).

Community factors. A strong relationship exists between poverty and risk of adverse child outcomes, including low academic skills at kindergarten entry (Schulman & Barnett, 2005). Data from the Early Childhood Longitudinal Study – Kindergarten Cohort (ECLS-K) of 1998-1999 showed that poverty was negatively related to literacy development from kindergarten to first grade (Kaplan & Walpole, 2005) and overall academic abilities (West et al., 2000; Zill & West, 2001). Children in poverty are three times more likely than those not in poverty to be born to an unmarried teenager and nearly seven times as likely to experience abuse and neglect (Shonkoff & Phillips, 2000). Poverty increases the chances that a child will be born at low birth weight, develop learning disabilities, be retained a grade in school or drop out of school altogether. Being poor not only exacerbates negative birth outcomes, but also places teen mothers at a disadvantage in terms of continuing their own education, which perpetuates a cycle of poverty for both mother and child (SmithBattle, 2007). Besides the obvious connection with poverty, parental unemployment is a school readiness risk factor for several reasons. Research suggests income insecurity contributes to low parental self-efficacy and decreased parental involvement, both factors that impact child-parent interactions and subsequently child development (Pelletier & Brent, 2002).

Another risk factor associated with low socioeconomic status is lead poisoning (Sargent et al., 1995), which has been found to have long-term detrimental effects on child development and healthy physical and behavioral functioning (Lane et al., 2008). Lead poisoning is considered to be a community factor due to concentrations of homes with lead paint in low-income areas (Sargent et al., 1995). Other factors found to be related to poor school readiness but not identified by the NSRII or ECCS include homelessness, limited English skills and parents who are migrant workers. In addition to negatively affecting reading and math performance in the third grade, homelessness has been found to be one of the strongest predictors of school absenteeism, suspensions and poor social skills (Fantuzzo et al., 2014; Rouse & Fantuzzo, 2009). As language development is a strong indicator of early reading skills, children living in homes and communities where little to no English is spoken are at a greater disadvantage for language development than children in English-fluent homes (Hair, Halle, Terry-Humen, Lavelle, & Calkins, 2006; West et al., 2000; Zill & West, 2001). This effect is even greater for children of migrant parents. Among all major demographic groups, migrant workers are recognized as the least educated, with most speaking little to no English (Nevárez-La Torre, 2012). Poverty is endemic among migrant families (Mathur & Parameswaran, 2011), whose children lack continuity of schooling and are often significantly behind in academic development (Green, 2003).

Health factors. Considerable research also points to the effects of several health-related issues, such as low birth weight, behind on vaccinations, and tobacco use during pregnancy, on school readiness. Poor birth outcomes, which result from inhibited in utero development of an infant's brain, can have significant negative impacts on a child's potential (NEGP, 1997; Thompson et al., 2009). For example, children with lower than normal birth weights or who were born pre-term are likely to have a learning or other type of disability at the start of kindergarten (Hair et al., 2006; High, 2008). Low birth weight, which can lead to permanent restrictions in physical and cognitive development (Kramer, 2003), contributes to racial and ethnic disparities in educational attainment, and can reduce future earnings potential (Behrman & Rosenzweig, 2004; Fiscella & Kitzman, 2009). Adequate prenatal care is essential for reducing the risk of poor birth outcomes and is as necessary to preparing a child for school as high-quality early childhood education (High, 2008).

Tobacco use during pregnancy and second-hand smoke exposure have been noted as among the primary contributors to negative birth outcomes, such as restricted fetal growth, preterm births and infant mortality (Bloch et al., 2010; Kramer, Séguin, Lydon, & Goulet, 2001; Tong et al., 2013). Between one quarter and one third of sudden infant deaths are attributable to prenatal cigarette use (Tong et al., 2013). Moreover, research suggests that smoking during pregnancy may mediate some of the relationship between low maternal education level and low birth weight (Finch, 2003). According to data from the Pregnancy Risk Assessment Monitoring System (PRAMS), Oklahoma had one of the highest rates of smoking during the last trimester of pregnancy (Tong et al., 2013).

Poor health, whether physical or emotional, can affect school readiness and subsequent academic performance in many ways. Students with frequent illnesses may be chronically absent; while those with behavioral problems, such as attention deficit and hyperactivity (ADHD) disorder, can significantly disrupt learning for themselves and even their peers. Children who are chronically absent due to a health condition are more likely to struggle with school work, be retained a grade in school, or drop out of school altogether than even the chronically absent without health issues (Klerman, 1988). Access to health care is a critical factor in reducing health-related risks. Research on the effects of enrollment in insurance programs designed for low-income populations showed that previously uninsured children received more frequent and timely medical care than a matched control group not continuously enrolled in the programs (Lave et al., 1998).

Childhood vaccinations are important to maintaining the health of children and reducing the risk of cognitive impairments that stem from preventable diseases, such as measles (Bloom, Canning, & Weston, 2005; *The National Education Goals report: Building a nation of learners*, 1999; Robinson, Sepe, & Lin, 1993). Bloom et al. (2005) found that vaccinations were positively related to IQ and language and mathematics performance among young children. Several studies have identified factors that contribute to under-vaccination, which include race/ethnicity, low maternal education, single-mothers, poverty, and birth outside the United States (Findley, Irigoyen, & Schulman, 1999; Luman, McCauley, Shefer, & Chu, 2003).

Racial/ethnic factors. Being of a racial/ethnic minority group increases the likelihood of experiencing multiple school readiness risk factors (West et al., 2000; Zill & West, 2001). Evidence shows that nearly 75% of Black and Hispanic children experience one or more risk factors, compared to 29% of White children (Zill & West, 2001). Considerably more Black and Hispanic kindergarteners live in poverty and have higher rates of low maternal education, teen mothers and single parents (Ducan & Magnuson, 2005). Regarding health factors, Black and Hispanic females have higher rates of teen pregnancies than Whites (Chandra et al., 2002; Hamilton et al., 2012; Martin et al., 2012); and Black mothers are twice as likely as White mothers to have low birth weight infants (Martin et al., 2012). Of children in foster care, Black (Ards, Myers Jr, Malkis, & Zhou, 2003) and American Indian/Alaskan Native children (Carter, 2010) are overrepresented.

American Indian children are least represented in early childhood education programs (Saluja, Early, & Clifford, 2002), followed by Hispanic children (Laosa & Ainsworth, 2007). Black and Hispanic children have been found to have lower social-emotional, cognitive, and language skills compared to White children (Carneiro & Heckman, 2003). American Indian children are overrepresented in learning disability status (Hosp & Reschly, 2004), and nearly 40% of Hispanic fourth graders are English language learners (National Center for Education Statistics [NCES], 2011). Research also shows that minority children have lower scores in grade-level mathematics and reading tests and higher rates of dropping out of school than their Caucasian peers (NCES, 2009; NCES, 2011).

Chapter Summary

This literature review examined the purposes and key features of indexes and examined methodological issues surrounding their construction. It continued by examining the construct of school readiness and discussing the transactional/ecological model of child development and the cumulative risk model. The use of both in this study provides a solid foundation for examining the characteristics of a school readiness risk index with respect to its indicators. Finally, this chapter highlighted several school readiness risk factors discussed frequently in the literature and recommended for regular monitoring by national indicator efforts. Several of these indicators will be used in this study.

CHAPTER III

METHODOLOGY

The purpose of this study was to examine the sensitivity of the Oklahoma School Readiness Risk Index (OK SRRI) (Lazarte Alcalá & Schumacher, 2014) to changes in the set of indicators used in its construction. The methods used in this study derived from several sources on index construction and sensitivity analysis, particularly those focused on the impacts of indicator selection (Helton et al., 1995; Houweling et al., 2003; Jones & Andrey, 2007; Saisana, 2008; Saisana & Munda, 2008; Scaglion & Condon, 1980; Schmidtlein et al., 2008). This study responded to one of the last steps of index construction, in which alternative methodological approaches are considered and analyzed for the extent of uncertainty they create in outcome rankings or scores (Nardo et al., 2008).

Research Questions

This study was guided by the following research questions, which are similar to those used by Saisana (2008) in assessing the robustness of the CLI, which measures lifelong learning across Canadian regions. The CLI assessment examined three key questions: the extent to which the index was "internally sound and robust with respect to its applications," could "withstand validation" using alternate measures of learning, and was sensitive to differing methods of construction, including varying indicator sets (Saisana, 2008, p. 2).

The research questions addressed by this study are listed below.

1. What is the relationship among scores on the overall index, the domains and individual indicators?
2. What is the impact of the indicators and domains on the overall index? In other words, do one or more indicators or domains dominate the index?
3. What is the relative effect of individual indicators on outcome rankings? In other words, how stable are index rankings when individual indicators or domains are removed from the index?
4. To what extent do changes to the indicators and domains affect associations of index rankings with a proxy outcome of school readiness risk?

Two indexes of county-level characteristics for the state of Oklahoma were analyzed for this study. The first was the original index as published by OKDHS (Lazarte Alcalá & Schumacher, 2014). The second was an alternate index derived from risk indicators used in the initial selection of indicators for the original index (Lazarte Alcalá et al., 2013) and organized into different domains. The indicator sets for both indexes represented a combination of variables from Table 3. The intent of this study was not to suggest that one indicator set is necessarily better than the other, but instead to assess the differences in rankings that might exist with the use of different sets representing different domains. All methods were conducted twice, once using the original index and again using the alternate index.

Data and Sources

This section describes the variables and sources of data used in this study. First, data used for risk indicators is discussed, which is followed by a description of the variable used as a proxy indicator of the consequence of starting school unready to learn.

Risk Indicators

The indicators considered for inclusion in the alternate index came from those considered for the original index with a few exceptions. First, the indicators of kindergarten retention and children in pre-kindergarten and kindergarten with Individual Education Plans were excluded from consideration for the alternate index as these indicators are considered a reflection, or effect, of school readiness risk rather than a cause. In addition, the indicator of prenatal care was modified from the definition used during construction of the original index of no prenatal care to include late prenatal care in order to correspond with NSRII recommendations. This change did not affect analysis of the original index as prenatal care was not included.

Indicator selection for the alternate index was guided by the theoretical framework of the transactional/ecological and cumulative risk models, the NSRII domains related to family, community and health factors, and the results of descriptive statistics, which revealed the extent to which indicators met necessary requirements for coverage and variability. With the exceptions noted above, indicators considered for inclusion in the original index are listed in Table 3. Indicators are organized according to the three NSRII domains noted above, which were used for the alternate index. Although race is technically considered a covariate, a type of indicator not ordinarily used in formative composite models (Bollen & Bauldry, 2011), race/ethnicity variables were included in the analysis because they were included in the original index and therefore required assessment as part of the indicator set.

This study used the same data examined in the process of selecting indicators for the original index (Lazarte Alcalá et al., 2013), as well as updated data used for the second publication of index findings (Lazarte Alcalá & Schumacher, 2014). The exception was the indicator of African American/Black race, which was not included in the original index and therefore not updated. For this study, data for this indicator were updated to match the timeframe of other race/ethnicity variables included in the original index and updated for the 2014 publication. Because data timeframes were restricted by what was most recently available from the source and the updating of only those indicators used in the original index, timeframes vary and range from 2007 to 2012.

The unit of analysis was counties. Data came from the following government sources: OKDHS, Oklahoma State Department of Education (OSDE), Oklahoma State Department of Health (OSDH), and the American Community Survey of the U.S. Census. Data were obtained either by contacting state agencies directly or from datasets made publicly available through online databases. It is important to note that data were not random but rather represented a census of all 77 Oklahoma counties.

Table 3

Indicators and Variables Assessed for Original and Alternate Indexes

Indicator	Variable
Ready Families	
Low maternal education*	Percent of live births to mothers without a high school diploma or GED of all maternal education levels (2008 to 2009)
Young maternal age*	Percent of live births to mothers ages 10 to 19 of all reported ages (2011 to 2012)
Abuse/neglect*	Percent of children under age 6 in poverty with substantiated cases of abuse and neglect (state fiscal year 2012)
Foster care*	Percent of children under age 6 in poverty in state protective custody (state fiscal year 2012)
Single parents*	Percent of children under age 6 living with single parents (2007 to 2011)
Ready Communities	
Poverty*	Percent of children under age 6 living under 100% of the federal poverty level (2007 to 2011)
Homeless children	Percent of children in pre-kindergarten and kindergarten defined as homeless by the McKinney-Vento Homeless Assistance Act (2009-2010)
English-language learners*	Percent of children in pre-kindergarten and kindergarten who are learning English (academic year 2011-2012)
Migrant children*	Percent of children ages 3 to 5 served by federal Migrant Education Program (academic year 2009-2010)
Ready Services - Health	
Vaccination rate	Percent of children ages 13 to 35 months who are behind the current vaccination schedule as recommended by the Advisory Committee on Immunization Practices, the American Association of Pediatrics, and the American Academy of Family Physicians (2011)
Low birth weight	Percent of infants born in 2012 weighing less than 2,500 grams (5.5 pounds) (2012)
Prenatal care	Percent of infants born to mothers who had no or late (third trimester) prenatal care (2012)
Tobacco use while pregnant	Percent of infants born to mothers who used tobacco during pregnancy (2012)
Race/Ethnicity	
Hispanic/Latino*	Percent of children under age 5 who are of Hispanic/Latino ethnicity (2007 to 2011)
American Indian*	Percent of children under age 5 who are of American Indian/Alaska Native race (2007 to 2011)
Black	Percent of children under age 5 who are of African American/Black race (2007 to 2011)

Note: *Items included in original index.

Data related to maternal characteristics, pregnancies and birth outcomes came from the vital statistics section of the OSDH web-based query system, Oklahoma Statistics on Health Available for Everyone (OK2SHARE), which allows users to locate data directly without requesting access. Demographic data related to children and mothers are reviewed by the National Center for Health Statistics before final publication (Lazarte Alcalá et al., 2013). Data on characteristics related to poverty, single parent families and race/ethnicity came from the annual American Community Survey (ACS), a product of the U.S. Census. The ACS is an annual survey designed to gather information related to basic demographic variables (e.g., age, sex, race), family structures and income, education and other data and is designed to aid in decision making related to infrastructure and services at the federal, state and local levels.

All remaining data were requested directly from state agencies by OKDHS (Lazarte Alcalá et al., 2013). Data on child abuse/neglect and children in foster care came from the OKDHS “KIDS” system, a statewide centralized child welfare information system that records data on abuse/neglect, foster care, adoptions and related variables. Data on child homelessness, English-language learners, migrants and learning disabilities were requested from OSDE and came from the agency’s student information tracking and reporting system, known as “the Wave,” which maintains federally required student demographic, enrollment, teacher and course data. Child vaccination data were requested from OSDH and came from the Oklahoma State Immunization Information System (OSIIS), a registry that collects and maintains current vaccination records for all Oklahomans of all ages. Although the OSIIS is the only source for vaccination data across the state, not all clinics report to it.

Proxy Outcome Variable: Literacy Nonproficiency

An additional variable was included in this study to respond to question 4 regarding the extent to which changes to the indicators and domains altered relationships with a variable theorized to reflect the outcome of being unready for school. To this end, this study used the county-level percent of kindergartners scoring below benchmark on one of three kindergarten-entry pre-literacy screening assessments approved for use during the 2012-2013 academic year. This year was selected as it was the last year that as few as three instruments were approved for use by the State Board of Education. Data were requested from OSDE by OKDHS researchers. The three literacy assessments used during this timeframe were the Dynamic Indicators of Basic Early Literacy Skills (DIBELS), Literacy First Phonological Awareness Skills Test (PAST) and phonics assessments, and the Basic Early Assessment of Reading (BEAR). Each Oklahoma public school district is required by law to administer at least one pre-literacy assessment (Reading Sufficiency Act of 1997, 2014). The availability of psychometric evidence for each instrument varies and is discussed below.

DIBELS. Created by researchers at the University of Oregon's Center on Teaching and Learning (Good & Kaminski, 2002), the DIBELS battery includes five main assessments measuring fundamental literacy skills, with four measures designed for use beginning at the kindergarten level (Shanahan, 2003). While more studies on the psychometric properties of DIBELS have been published compared to BEAR or Literacy First, the evidence is mixed and varies by specific DIBELS measure. In a review of 26 psychometric studies, Goffreda and DiPerna (2010) found the only consistent evidence of strong validity and reliability related to the DIBELS measure designed for use beyond kindergarten. Of those geared toward kindergarten, reliability testing included alternate forms and test-retest, with coefficients ranging from a low of .58 to a high of .97. Likewise, concurrent validity testing using a variety of instruments yielded a wide range of coefficients. Median values for the kindergarten-level assessments ranged from .33 to .58, with some measures performing better overall than others. For two measures, concurrent validity testing resulted in coefficients below .10. Similar evidence was found in studies of DIBELS predictive validity, with coefficients ranging from as low as .15 to a high of .93 (Goffreda & DiPerna, 2010).

BEAR. A product of Riverside Publishing, BEAR is a criterion-referenced comprehensive battery of reading assessments for children from kindergarten through third grade with four assessments available for each grade level (Gratz, 2002; Rathvon, 2004). The Initial-Skills subtest is designed to be administered at the start of the school year. The technical manual reports the results of reliability and validity studies using students from 37 schools nationwide during the 2001-2002 academic year (Gratz, 2002). For the paper-and-pencil version of the Initial Skills subtest, coefficients alpha ranged from .83 to .88 across all grade levels, while four-week test-retest studies yielded reliability estimates from .70 to .77. Similar coefficients were found for the computer-administered versions. Evidence of predictive validity was presented through evaluations of the relationship between Initial Skills component scores and both skill-specific and summative subtests. Correlations ranged from .38 to .61 for kindergarten and first grade tests (Gratz, 2002).

Literacy First. The Literacy First assessments are part of an instructional framework of Catapult Learning that includes reading curricula specific to early childhood, elementary school, and middle and high school content areas ("Literacy First framework," 2014). A search of the peer-reviewed literature found no published studies of the psychometric properties of any Literacy First assessments. The only evidence found for the Literacy First PAST or phonics assessments comes from a paper on the company's web site ("Comparison of Literacy First and DIBELS assessments," 2005). According to this document, reliability for both the PAST and phonics assessment was measured using Cronbach's coefficient alpha. For the various PAST measures, values ranged from .84 to .95, with an overall coefficient of .96. For the various phonics assessments, it was reported that all coefficients either approached or exceeded .80 and were similar to the Woodcock Reading Mastery Word Identification subtest. Concurrent validity of both PAST and the phonics assessments was determined to be acceptable, with correlations with the Woodcock Word Identification subtest of .71 for the PAST composite score and at least .77 for the phonics measures.

Procedures

This section begins with a discussion of the methods used to develop the original OK SRRI and derive index scores and continues with a description of the statistical methods used to answer the research questions.

Oklahoma School Readiness Risk Index Development Process

The OK SRRI originated as a partnership between OKDHS and the Oklahoma Partnership for School Readiness, also known as Smart Start Oklahoma, in response to a statewide initiative to increase access to quality early childhood education and child care programs. Planned for annual updating, the index is intended to be used in tandem with measures of access to and availability of early childhood programs to pinpoint counties of greatest need and inform policy and resource allocation decisions at the state level. Several steps were used to construct the index.

First, a comprehensive review of empirical studies related to school readiness was conducted to identify the most salient indicators of school readiness risk. Due to limitations related to data availability at the county level, 18 indicators were ultimately selected for consideration. Indicators were conceptually organized into the domains of infant/child, maternal and family factors and included indicators such as poverty, minority race/ethnicity, family structure, abuse/neglect and various health indicators. Data were collected at the county level.

Statistical methods were used to select final indicators for the index, with the goal of identifying statistically significant predictors of being unready for school (Lazarte Alcalá et al., 2013). First, PCA was used as a data reduction technique, with the unit of analysis all of the state's 77 counties, or the population of counties in the state. This resulted in 4 components comprised of 12 variables, and 6 variables that did not strongly correlate with any single component. Next, OLS multiple regression analysis was conducted to identify variables significantly associated with school readiness. Component scores and data for the six individual variables that failed to load on a single component were used as predictor variables in two regression models, one using percent of third-grade students scoring below the satisfactory level in reading on the Oklahoma Core

Curriculum Test as the criterion variable, and the second using percent of third-grade students scoring below the satisfactory level in mathematics. Both variables were measured at the county level. Only the reading model had a significant F statistic ($p < .05$), and one outlier was detected. The adjusted R^2 was 0.16 with the outlier and 0.24 after the outlier was removed. Its removal did not affect the significance of individual coefficients, of which three of the four components and one individual indicator, migrant children, were significant ($p < .05$). As a result, the final 10 indicators selected for the index consisted of the 9 indicators that comprised the three significant components and the migrant indicator. Table 4 lists these components as named in the original index and their associated indicators (Lazarte Alcalá et al., 2013; Lazarte Alcalá & Schumacher, 2014). Each component was used as an index domain, with scores calculated for each domain following the method used for overall index scores.

Table 4

Oklahoma School Readiness Risk Index Components and Indicators

Component/Domain	Indicator
Hispanic Background	Hispanic/Latino ethnicity
	English-language learners
	Low maternal education
	Migrant children*
Family Structure and Economic Distress	Poverty
	Young maternal age
	Single-parent family
	American Indian/Alaska Native race
Children in Child Welfare	Abuse/neglect
	Foster care

Note: *Although not reported to load on this component, this indicator was conceptually organized with this component but not included in calculations of index domain scores for the published index. For convenience, it is included in the Hispanic Background domain for this study.

The formula used to derive overall scores for each county was changed from the first publication of the index in 2013 to its subsequent publication in 2014. For the 2013 index, scores were derived by assigning each county a value of 0 (low risk) to 3 (high risk) when compared to the national average for each indicator. Those below the national average were assigned a value of 1 if observations were above 0% and a value of 0 otherwise. Next, for each indicator, the arithmetic median was calculated for all remaining county data above the national average. Counties below the median were assigned a value of 2, and those above it were assigned a value of 3. The arithmetic mean was calculated for each county across all 10 indicators, with scores ranging from 1.0 to 2.5, and counties were categorized into three risk groups of low (1.0 to 1.4), medium (1.5 to 2.4) and high (2.5).

This approach resulted in numerous counties having the same index score, which prevented meaningful rankings, and little variability in risk groupings as only one county was ranked as high risk. To increase variability in the scores, the formula was changed for the next publication (Lazarte Alcalá & Schumacher, 2014). Instead of assigning county scores for each individual indicator based on the national average, county-level observations were transformed into *z*-scores using the mean and standard deviation of individual variable distributions across Oklahoma counties, with overall index scores derived using the arithmetic mean *z*-score across the 10 indicators. There were no weights assigned to indicators, so all indicators were considered equally important. Scores were arranged in descending order and counties ranked from 1 (highest risk) to 77 (lowest risk). Quartile distributions were then used to classify counties into four groups of high, high-medium, medium-low and low risk. This resulted in increased variability and the ability to rank counties without numerous ties. Domain scores were computed and counties classified on each domain the same way. Although the 2014 index used updated data for some indicators as available and changed the operational definition of poverty from 200% to 100% of the federal poverty level, the indicators were the same as those selected on the basis of statistical analysis of data used for the 2013 index.

Data Analysis Plan

Data were analyzed for this study using IBM SPSS statistics software version 22.0 and Microsoft Excel 2013. Methods followed recommended steps for 1) selecting and evaluating the relative importance of index indicators, and 2) linking with a variable thought to represent an outcome of the construct being measured (Nardo et al., 2008; Saisana, 2008; Saisana & Munda, 2008). Answering the research questions first required selecting indicators for use in the alternate index. Essentially the alternate index was a modified version of the original index with a different domain structure. Based on the criteria to include an indicator, described below, it was expected that the alternate index would consist of many of the indicators selected for the original index, in addition to some indicators considered for inclusion in the original index but ultimately excluded. These indicators were largely health related.

Selecting indicators for the alternate index required consideration of their analytical soundness, measurability, coverage, relevance and relationships to each other, particularly in terms of representing the theorized domains. Indicator characteristics were evaluated by examining descriptive statistics for all variables from Table 3, with particular attention paid to means, standard deviations, range of observations, missing values, correlations and the presence of outliers (e.g., 1.5 x interquartile range (IQR)). To be included in the alternate index, variables must have had observations for most of the state's 77 counties and at least a moderate amount of variability and range to detect differences among counties. This follows the approach used to select indicators for the Illinois school readiness risk index. Among several variables considered, an indicator was excluded from the state's index if it failed to impact county variation, was highly correlated with another variable, or contributed little to the total score (Thomas et al., 2012). The first two criteria were used to select indicators for use in the alternate index, while the third research question addressed indicator impact. Scores on the alternate index were computed using the same methods of z-score transformations and aggregation used for the original index published in 2014.

Research question 1. *What is the relationship among scores on the overall index, the domains and individual indicators?* To answer question 1, the approach used by Saisana (2008) was employed for both the original and alternate indexes. Pearson's *r* product-moment correlations were examined to evaluate the relationships between overall index scores and scores on the domains and individual indicators. Although it has been recommended that an individual indicator with a weak correlation with overall score should be removed, developers are cautioned against assuming that strong correlations mean an indicator should remain (Saisana, 2008). The decision to include or exclude an indicator should be driven by its adequacy and quality as a variable. An indicator should be considered for the extent to which it is known to be associated with the construct and is problematic across a particular area, such as a state or region.

Since higher scores on all dimensions of the original and alternate indexes indicate greater risk, these dimensions should be positively correlated with each other and with index scores (Saisana, 2013, Nov. 12). It is desirable for the relationships among the dimensions to vary, as this suggests the domains account for different aspects of the construct, yet partially overlap and are therefore inseparable (Saisana, 2008).

Research question 2. *What is the impact of the indicators and domains on the overall index? In other words, do one or more indicators or domains dominate the index?*

Commonality analysis was used to answer question 2. A method of partitioning variance, the goal of commonality analysis is to determine the proportion of variance in a dependent variable uniquely explained by an independent variable and the proportion shared in combination with other variables (Pedhazur, 1997). Dividing the total effect of a predictor by R^2 yields its squared structure coefficient, which indicates its variance-accounted-for effect size. This coefficient may be driven by a variable's unique effect or the effect it shares in common with other predictors (Nimon, 2010). Unique effects represent the change in R^2 that would occur if a variable was entered last in a hierarchical regression model, making the unique contribution a squared semipartial correlation between the independent and dependent variable when the effects of other independent variables are partialed out (Pedhazur, 1997).

By identifying and accounting for the effects of suppression and multicollinearity, commonality analysis overcomes the problems of using beta weights or stepwise methods to assess the relative importance of predictors (Reichwein Zientek & Thompson, 2006). For index construction, commonality analysis can be used as a means of identifying whether any indicators dominate the index, and the extent to which indicators share effects. A large unique effect relative to other indicators suggests an indicator dominates an index, whereas large shared effects among most indicators suggests indicators may not tap a construct's multidimensional nature.

Commonality analysis was conducted four times, twice to examine the indicator sets for both the original and alternate indexes and twice to examine domains for both indexes. For each index, the overall score served as the dependent variable, with indicator and domain scores *z*-scores used as independent variables. This approach provided insight into whether results were being primarily driven by particular indicators or domains and identified the relative importance of each. It is desirable that results show all indicators contribute some impact on index scores, and that the impact is not dominated by a small number of indicators. The impact of individual domains should also be distributed across domains rather than the index being dominated by one or two domains (Saisana, 2008).

Research question 3. *What is the relative effect of individual indicators and domains on outcome rankings? In other words, how stable are index rankings when individual indicators or domains are removed from the index?* Another way of demonstrating the relative impact of individual indicators is to examine how much rankings change when indicators are removed from the index. Question 3 was answered by calculating various scenarios of indicator sets, each time leaving out one indicator or domain, and comparing county rankings on these reduced sets to their full index ranking. For examining domains, the full indexes were recalculated three times (corresponding to three domains for each index), each time excluding the individual indicators that comprised each domain. This corresponds to the method used to construct the original index of aggregating individual indicator scores rather than domain scores. This process was conducted for each set of indicators used in the original and alternate indexes. Scores were computed for both indexes as well as for their associated reduced scenarios and ranks assigned using percentiles, with higher percentiles representing greater risk. For example, a county at the 70^{th} percentile is considered at higher risk than 70% of the counties in the dataset. Differences in ranks between the reduced and full original and alternate indexes were calculated, with the full indexes serving as the reference point for each respective set.

Indicator and domain impacts were examined at two levels. The first assessed the impact of indicators and domains overall, and the second evaluated impacts at the county level. To assess the magnitude of overall impacts, percentile differences were converted to absolute values. The median and maximum absolute shifts in rank across the reduced scenarios, as well as the number of counties with moderate to significant shifts were reported. As there are no established thresholds from which to evaluate shifts in rankings and make judgements regarding their importance, a cutpoint of 15 percentiles was selected as a meaningful change based on findings reported in other studies. A common lower bound threshold for significant rank changes reported by several researchers is approximately 20 percentiles (Jones & Andrey, 2007; Mather & Dupuis, 2012; Schmidtlein et al., 2008), although others maintain that shifts less than 25 (Saisana, 2008) or 30 percentiles are relatively minor (Houweling et al., 2003). For this study, an increase or decrease in rank of at least 20 percentiles was considered significant. To provide a broader perspective of county-level impact, shifts in rank of 15 to 19 percentiles were reported and considered moderate.

To examine county-level impact, maximum rank changes, in absolute values, were plotted relative to each county's full index ranking for counties with one or more shifts in rank of at least 15 percentiles across indicator- and domain-reduced scenarios. Comparing to the full index rank provided greater insight into the extent to which shifts affected interpretation of the rankings. For example, a shift could result in a county originally classified as moderate risk increasing to a classification of high risk or decreasing to a classification of relatively low risk depending on its full index ranking and magnitude and direction of the effect. The amount by which counties with meaningful shifts increased and/or decreased in rank on indicator- and domain-reduced scenarios was also reported. This approach identified which risk factors were the most influential for a particular county's ranking.

The analyses for question 3 concluded with an examination of the association between the two indexes using a Spearman's rank order correlation coefficient. In addition, counties with meaningful differences in ranks between the full original and alternate indexes were identified. For counties with absolute shifts of at least 15 percentiles, ranks on the alternate index were plotted relative to original index rankings. Changes to risk groupings between the full and alternate index were also assessed. Ranks on the alternate index were used to classify counties according to four categories of high to low risk following the methodology used in the original index. Counties with rank changes great enough to move them to a higher or lower risk category were reported, as well as the actual rank increase or decrease from the alternate to the original index. This followed the approach used by Schmidtlein et al. (2008) to compare the full Social Vulnerability Index to an alternate index with a reduced number of variables.

Research question 4. *To what extent do changes to the indicators and domains affect associations of index rankings with a proxy outcome of school readiness risk?* The scenarios used for question 3 were again used for question 4, which responded to the recommended step of linking indexes with other measures of the same phenomenon. The percent of Oklahoma kindergartners scoring below proficiency on one of three kindergarten-entry pre-literacy screening assessments was used as a proxy for starting school unready to learn. Using percentiles, counties were ranked from highest to lowest rates of children scoring below proficiency. The relationship between literacy rankings and rankings on the full and indicator- and domain-reduced indexes were computed using Spearman rank correlation coefficients. This is similar to the method used for CLI validation testing, which consisted of correlating scores on the overall index and multiple scenarios with scores on the Economic and Social Well-Being index and associated individual indicators, considered outcomes of lifelong learning (Saisana, 2008).

Chapter Summary

This chapter discussed the context of the study and the development of the OK SRRI, which served as the sample index for this study. Further, this chapter discussed the data to be used and their sources. It listed the four research questions that guided this study and provided detail on the specific analytic methods that were used to answer these questions.

CHAPTER IV

FINDINGS

The purpose of this study was to assess the extent to which indicator selection affected county rankings on an index of risk for poor school readiness. This study addressed the following research questions.

1. What is the relationship among scores on the overall index, the domains and individual indicators?
2. What is the impact of the indicators and domains on the overall index? In other words, do one or more indicators or domains dominate the index?
3. What is the relative effect of individual indicators and domains on outcome rankings? In other words, how stable are index rankings when individual indicators or domains are removed from the index?
4. To what extent do changes to the indicators and domains affect associations of index rankings with a proxy outcome of school readiness risk?

Indicator Characteristics

The first step in the analysis was to examine zero-order correlations and descriptive statistics of indicators used in the original index and those considered for the alternate index. Data were analyzed for missing data and the presence of outliers, with no missing data found. No outliers were found for the indicators of single parent and young maternal age or for the proxy outcome measure of percent of entering kindergartners with literacy skills below proficiency. The indicators of behind on vaccinations, low birth weight, foster care and American Indian race each contained one outlier county. Late or no prenatal care and tobacco use during pregnancy each had two outlier counties. The indicator of abuse/neglect contained three outlier counties, as did low maternal education. Poverty contained five outliers, and seven counties were outliers on the Hispanic ethnicity indicator. Of these counties, five were also outliers on the English-language learner (ELL) indicator, which included five additional outlier counties.

The African American/Black race indicator had 10 outlier counties, and 29 counties (38%) with no rates of Black children under age 5. The indicator of homeless children had rates for 17% of counties, while the migrant indicator had rates for 8% of counties, making all of these counties outliers. With the exception of the migrant indicator, used in the original index, these indicators were excluded from further analysis due to limited county coverage. In total, 23 counties were outliers on at least one of the remaining 14 indicators as shown in Table 5. Although discussion of the final selection of indicators for the alternate index follows analysis of correlations, indicators used in each index are denoted here for convenience and ordered by index (e.g., the first 10 indicators were used in the original index, and the last 11 in the alternate index). Table 3 in Chapter 3 provides a description of indicators, and Table 6 on the next page lists indicator labels.

Table 5

Outlier Counties by Indicator

	Indicator													
County	AI[1]	MIGRANT[1]	HISPANIC[1]	ELL[1,2]	POVERTY[1,2]	SPARENT[1,2]	MAGE[1,2]	MEDUC[1,2]	ABUSE[1,2]	FOSTER[1,2]	VACCINE[2]	BIRTHWT[2]	PRENATAL[2]	TOBACCO[2]
Adair	✓			✓				✓						
Beaver		✓	✓	✓										✓
Caddo		✓												
Cherokee		✓												
Cimarron			✓		✓									
Coal									✓					
Dewey													✓	
Ellis											✓			
Greer									✓	✓				
Harmon			✓	✓	✓			✓						
Harper				✓										
Hughes					✓									
Jackson		✓	✓											
Kingfisher				✓										
Marshall			✓	✓										
Oklahoma				✓										
Ottawa														✓
Pushmataha					✓									
Seminole									✓					
Sequoyah													✓	
Texas		✓	✓	✓				✓						
Tillman		✓	✓	✓	✓							✓		
Tulsa				✓										

Note: N = 77. [1]Indicators used in original index. [2]Indicators selected for alternate index.

Table 6 lists descriptive statistics for the 14 indicators used in either the original or alternate index or both, as well as the variable of literacy nonproficiency, considered a consequence of starting school unprepared. The indicators of abuse/neglect, foster care and migrant all had mean rates of less than 3%. These indicators, in addition to low birth weight, all had standard deviations of no more than 2% and ranges of 1% to 9%. This suggests these indicators may have little impact on county variation.

Table 6

Descriptive Statistics for Indicators and Proxy Outcome Variable

Indicators	Label	*M*	*SD*	Min	Max	Range
American Indian race[1]	AI	10.87	8.96	0	41.5	41.5
Migrant education[1]	MIGRANT	0.04	0.17	0	1.2	1.2
Hispanic ethnicity[1]	HISPANIC	12.83	9.52	3.3	58.6	55.3
English-language learner[1,2]	ELL	6.54	9.99	0	57	57
Poverty[1,2]	POVERTY	28.18	10.27	11	54.9	43.9
Single parent family[1,2]	SPARENT	28.69	9.2	7.6	50.9	43.3
Young maternal age[1,2]	MAGE	13.54	3.4	6.8	22	15.2
Low maternal education[1,2]	MEDUC	20.95	6.47	8.8	48.2	39.4
Abuse and neglect[1,2]	ABUSE	1.95	1.22	0	8.2	8.2
Foster care[1,2]	FOSTER	2.47	1.51	0	7.9	7.9
Late or no vaccines[2]	VACCINE	24.15	4.88	14.5	39	24.5
Low birth weight[2]	BIRTHWT	7.38	1.74	2.5	11.4	8.9
Late or no prenatal care[2]	PRENATAL	7.84	3.51	0	22.2	22.2
Tobacco use while pregnant[2]	TOBACCO	18.42	5.72	7.8	35.1	27.3
Below literacy proficiency		41.63	16.09	4.8	86	39.4

Note: N = 77. [1]Indicators used in original index. [2]Indicators selected for alternate index.

Correlations among indicators were also examined. Even though the data represent a population, for which the correlation coefficient is represented by ρ, for convenience the common nomenclature of r, the coefficient for a sample, will be used for all correlational analyses conducted for this study. Table 7 lists correlations among the 14 index indicators. For convenience, the correlation matrix for the original and the alternate index is indicated.

Table 7

Zero-order Correlations among Indicators

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. AI[1]														
2. MIGRANT[1]	-.20													
3. HISPANIC[1]	-.36	.47												
4. ELL[1,2]	-.25	.33	.81											
5. POVERTY[1,2]	.28	.13	.09	.06										
6. SPARENT[1,2]	.23	.17	.06	.05	.72									
7. MAGE[1,2]	.26	.04	.08	-.05	.37	.32								
8. MEDUC[1,2]	.20	.15	.55	.60	.47	.39	.49							
9. ABUSE[1,2]	.05	-.21	-.09	-.25	.13	-.01	.41	.05						
10. FOSTER[1,2]	.10	-.25	-.13	-.29	.17	.07	.43	.07	.80					
11. VACCINE[2]	-.02	-.03	-.15	-.20	-.04	-.01	-.31	-.29	-.17	-.18				
12. BIRTHWT[2]	.05	-.18	-.06	-.13	-.06	.16	.11	.07	-.07	.06	.03			
13. PRENATAL[2]	.17	.12	.08	.06	.07	.08	.26	.24	.15	.16	-.04	-.22		
14. TOBACCO[2]	-.16	.08	-.03	-.10	-.19	-.18	.20	-.07	.14	.14	-.32	.03	-.16	

——— original index
- - - - - alternate index

Note: N = 77. [1]Indicators used in original index. [2]Indicators selected for alternate index.

In general, across the set, there were relatively few extremely high correlations, which means indicators represented the multidimensional nature of school readiness risk. Large correlations were found between the indicators of ELL and Hispanic (r = .81), abuse/neglect and foster care (r = .80), and poverty and single parent (r = .72). All of these indicators were used in the original index and all but one (Hispanic) in the alternate index. Remaining correlations across the entire set ranged from r = .60 between ELL and low maternal education to r = -.36 between American Indian and Hispanic. The original index had eight bivariate correlations of approximately r = .50 or higher. Excluding the Hispanic indicator from the alternate index removed three of these correlations, leaving this index with the same remaining five correlations of $r \geq .50$ as in the original index. These were the only correlations of this magnitude in the alternate index. The high correlation between Hispanic and ELL, and the desire for the alternate index to exclude indicators specific to race/ethnicity, suggested that ELL may be an effective proxy for the Hispanic indicator.

Ten indicators had negative correlations, which, in general, were small. Most indicators with negative correlations had only two to three such correlations. The exceptions were the indicators of vaccinations and tobacco use, which had five and six negative correlations, respectively. For the most part, correlations for the tobacco indicator were small ($r \leq -.19$). The exception is its moderate correlation with the vaccinations indicator ($r = -.32$), which also had a moderate association with young maternal age ($r = -.31$) and low maternal education ($r = -.29$). Although counterintuitive, these correlations make sense to some degree given the current anti-vaccination movement (Camargo & Grant, 2015). Mothers with higher levels of education may be more likely to read about and consider arguments for not vaccinating their children, whereas younger mothers and/or those of less education may be more likely to follow the advice of their doctor. As noted in Chapter 2, correlations of aggregated data may not follow the same patterns as what would be found with individual-level data. Therefore, negative correlations between indicators do not necessarily mean these indicators should be excluded from an index.

Based on the above information and in context of the theoretical framework for school readiness risk described in Chapter 2, the alternate index included all indicators except homelessness and migrant due to low county coverage, and race/ethnicity due to the desire to focus on indicators that can be addressed through public policy and interventions. Although descriptive statistics suggested the indicators of foster care and abuse/neglect might have little impact on overall outcome rankings, they were included in the alternate index as these are important issues for OKDHS, and nearly every county had at least some children in these situations. Similarly, despite concerns regarding low variability and negative correlations, the indicators of low birth weight, behind on vaccinations and tobacco use during pregnancy were included to allow for continued analysis of these indicators.

Table 8 lists indicators selected for the original and alternate indexes and indicates their associated domains.

Table 8

Indicators and Domains Used in the Original and Alternate Indexes

Indicator	Original index	Alternate index
AI	Family Structure/Economic Distress	-------
MIGRANT	Hispanic Background	-------
HISPANIC	Hispanic Background	-------
ELL	Hispanic Background	Ready Communities
POVERTY	Family Structure/Economic Distress	Ready Communities
SPARENT	Family Structure/Economic Distress	Ready Families
MAGE	Family Structure/Economic Distress	Ready Families
MEDUC	Hispanic Background	Ready Families
ABUSE	Children in Child Welfare	Ready Families
FOSTER	Children in Child Welfare	Ready Families
VACCINE	-------	Ready Services - Health
BIRTHWT	-------	Ready Services - Health
PRENATAL	-------	Ready Services - Health
TOBACCO	-------	Ready Services - Health

Research Question 1

The first research question for this study asked, *What is the relationship among scores on the overall index, the domains and individual indicators?* To answer this question for both the original and alternate indexes, Pearson's *r* product-moment correlations were used to examine the relationships between overall index scores and scores for the individual indicators and domains.

Indicators. Table 9 lists the correlations between indicators and the original index. All correlations were positive, which is desired since the index score is an exact linear combination of the indicators. There were no correlations between the overall original index and individual indicators that were of concern or unexpected. The largest correlation was with the indicator of low maternal education (r = .79), followed by poverty (r = .68), young maternal age (r = .67) and single parent (r = .60). The smallest correlations were with the indicators of migrant (r = .32) and American Indian (r = .26). The remaining indicators of Hispanic, ELL, abuse/neglect and foster care all had moderate correlations with the overall index score.

Table 9

Pearson Correlation Coefficients between Original Index and its Indicators

Hispanic Background		Family Structure/ Economic Distress		Children in Child Welfare	
HISPANIC	.50	POVERTY	.68	ABUSE	.38
ELL	.40	MAGE	.67	FOSTER	.39
MEDUC	.79	SPARENT	.60		
MIGRANT	.32	AI	.26		

Note: N = 77.

Table 10 lists the correlations between indicators and the alternate index, with the mean *z*-score as the overall index score. The largest correlation was with the indicator of young maternal age (*r* = .74), followed by low maternal education (*r* = .68), poverty (*r* = .61) and single parent (*r* = .59). The indicators of abuse/neglect, foster care and prenatal care were moderately correlated with the alternate index score. The smallest correlations were with the indicators of low birth weight (*r* = .22), ELL (*r* = .17), tobacco use (*r* = .11) and vaccinations (*r* = -.12).

Table 10

Pearson Correlation Coefficients between Alternate Index and its Indicators

Ready Families		Ready Communitiies		Ready Services - Health	
MEDUC	.68	POVERTY	.61	VACCINE	-.12
MAGE	.74	ELL	.17	BIRTHWT	.22
ABUSE	.49			PRENAT	.36
FOSTER	.55			TOBACCO	.11
SPARENT	.59				

Note: N = 77.

Indicators that are theorized to move in the same direction as the index should be positively correlated with index scores. Therefore, the negative correlation between the vaccinations indicator and overall score is of concern as this suggests high rates of children behind on vaccination schedule would reduce a county's school readiness risk score, even if to a small extent.

Of the seven indicators common to both indexes, there were some notable changes from the original to the alternate index. The correlation coefficient for the ELL indicator declined in magnitude by .23, while the coefficients for abuse/neglect and foster care increased by .12 and .16, respectively. The coefficients for young maternal age and low maternal education switched somewhat in terms of magnitude, with the coefficient for low maternal age dropping by .11 from the original to the alternate index and the coefficient for young maternal age increasing by .07.

Domains. Regarding domain scores, there should be positive associations with index scores and with each other. Relationships among the domains should vary as this indicates they represent the multidimensional construct of school readiness risk, and they should have some degree of correlation. This indicates the domains are tightly connected; yet by themselves they convey information relevant to particular issues of importance. Table 11 lists correlations between full original index scores and domain scores.

Table 11

Pearson Correlation Coefficients between Original Index and its Domains

	Hispanic Background	Family Structure/ Economic Distress	Children in Child Welfare
Overall score	.64	.77	.41
Hispanic Background		.15	-.19
Family Structure/Economic Distress			.25

Note: N = 77.

All correlations were relatively moderate, with the highest being the correlation with the Family Structure/Economic Distress domain (r = .77), followed by the association with the Hispanic Background domain (r = .64) and Children in Child Welfare (r = .41) domain. Regarding correlations between domains, the most associated were the domains of Family Structure/Economic Distress and Children in Child Welfare (r = .25). The Hispanic Background domain had very low correlations with the other domains. This is not necessarily problematic, except for the negative correlation with Children in Child Welfare (r = -.19). This suggests these two domains work against each other, even if to a small degree. This is likely due to the moderate negative correlations of the ELL indicator and to a lesser degree the Hispanic indicator with the two indicators of abuse/neglect and foster care that comprise the Children in Child Welfare domain.

As shown in Table 12, correlations of domain scores on the alternate index with overall scores ranged in magnitude from moderately to highly associated.

Table 12

Pearson Correlation Coefficients between Alternate Index and its Domains

	Ready Families	Ready Communities	Ready Services - Health
Overall score	.92	.56	.35
Ready Families		.40	.09
Ready Communities			-.26

Note: N = 77.

The alternate index was highly correlated with the domain of Ready Families (*r* = .92) and moderately associated with Ready Communities (*r* = .56) and Ready Services – Health (*r* = .35). While correlations of domains with overall index scores were less balanced in the alternate than in the original index, correlations between domains were more varied in the alternate index. The most associated were the Families and Communities domains (*r* = .40) and the least associated the Famillies and Health domains (*r* = .09). As noted above, these results are acceptable as they demonstrate the domains account for varying aspects of school readiness risk. There was one negative correlation between the domains of Communities and Health (*r* = -.26). Again, this suggests these domains work against each other, and is likely explained by the fact that both indicators that comprise the Communities domain (poverty and ELL) had small negative correlations with three of the four indicators that comprise the Health domain. The prenatal care indicator was the only indicator positively associated with poverty and ELL.

Research Question 2

The second research question asked, *What is the impact of the indicators and domains on the overall index? In other words, do one or more indicators or domains dominate the index?* To answer this question, commonality analysis was conducted using individual indicator and domain scores to determine the proportion of variance in index scores uniquely explained by an indicator and the proportion shared with other indicators.

Indicators. Table 13 presents the results of multiple regression and commonality analysis for the original index, with overall score as the dependent variable and indicators, ordered by magnitude of total effect, as independent variables. Because the overall score is comprised entirely of data from the indicators, the model R^2 is 1.0.

Table 13

Regression Results: Effect of Indicators on Original Index Scores

Indicator (x)	R	R^2	R^2_{adj}	β	p	*Unique*	*Common*	*Total*
	1.0	1.0	1.0					
MEDUC				.200	.000	.010	.617	.627
POVERTY				.201	.000	.017	.448	.464
MAGE				.199	.000	.019	.429	.448
SPARENT				.200	.000	.018	.346	.364
HISPANIC				.202	.000	.010	.235	.245
ELL				.198	.000	.009	.152	.161
FOSTER				.200	.000	.013	.142	.155
ABUSE				.200	.000	.014	.128	.142
MIGRANT				.200	.000	.026	.078	.105
AI				.202	.000	.026	.042	.068

Note: N = 77. *Unique* = x's unique effect. *Common* = ∑ x's common effects. *Total* = *Unique* + *Common.*

In this instance, it is clear that using beta weights alone, all approximately .20, would provide no information as to the impact of the indicators. The commonality analysis showed considerable common effects among the indicators and little unique effects, suggesting that no indicator by itself dominated the index. The indicators of migrant and American Indian contributed the largest unique effects, yet had the least in common with other indicators and contributed the lowest total effects. Five indicators (low maternal education, poverty, young maternal age, single parent and Hispanic) had common and total effects ranging, in descending order, from .62 to .24, while the indicators of ELL, foster care and abuse/neglect all had similar common and total effects ranging from .16 to .13.

Table 14 shows results for the alternate index. The model R^2 is again 1.0 since all indicators comprise the overall score.

Table 14

Regression Results: Effect of Indicators on Alternate Index Scores

Indicator (x)	R	R^2	R^2_{adj}	β	p	*Unique*	*Common*	*Total*
	1.0	1.0	1.0					
MAGE				.226	.000	.023	.517	.540
MEDUC				.227	.000	.014	.452	.466
POVERTY				.228	.000	.018	.353	.371
SPARENT				.228	.000	.022	.329	.351
FOSTER				.227	.000	.017	.286	.303
ABUSE				.228	.000	.018	.225	.242
PRENAT				.228	.000	.040	.093	.133
BIRTHWT				.226	.000	.037	.011	.049
ELL				.226	.000	.020	.008	.028
VACCINE				.227	.000	.039	-.025	.014
TOBACCO				.226	.000	.038	-.026	.012

Note: N = 77. *Unique* = x's unique effect. *Common* = ∑ x's common effects. *Total* = *Unique* + *Common.*

As with the original index, there was little unique effect of any indicator. Six indicators had common and total effects ranging from .54 to .23. Ordered by magnitude of effect, these were young maternal age, low maternal education, poverty, single parent, foster care and abuse/neglect. The first four indicators were the same indicators that made the largest contributions to original index scores. These indicators all had considerably more common than unique effects, while four indicators (low birth weight, ELL, vaccinations and tobacco use) all had very little common effects. Although unique effects for these indicators were higher than for the first six indicators, they were extremely small, with total effects of less than .05. The prenatal indicator had higher common than unique effect but had a small total effect (.13) relative to the six indicators that made the largest contributions. The negative shared effects of the vaccinations and tobacco use indicators, which reduced their total effects, were consistent with the numerous negative bivariate correlations associated with these indicators.

There are several notworthy differences between the alternate and original indexes. The distance between indicators that had the largest and second largest total effects on overall scores was .07 in the alernate index compared to .17 in the original index, suggesting less impact by a single indicator in the alternate index. The indicators of foster care and abuse/neglect had larger effects on the alternate index than on the original, making these indicators among the six most influential compared to being among the three least influential in the original index. The ELL indicator had considerably less impact on overall scores in the alternate than in the original index. ELL had a higher common than unique effect on the original index, whereas this pattern was reversed in the alternate index, with this indicator having very little in common with other indicators. This suggests the effect of ELL on original index scores was largely shared with the Hispanic indicator and to a lesser extent shared with the migrant indicator.

Domains. As shown in Table 15, commonality analysis of domain scores for the original index revealed a relative balance in terms of total effects for the domains of Family Structure/Economic Distress (.59) and Hispanic Background (.41). The domain of Children in Child Welfare had a considerably lower total effect (.17) on overall scores. Both the Family Structure/Economic Distress and Hispanic Background domains had approximately equal unique effects at .31 and .33, respectively. However, while the Family Structure/Economic Distress domain had a moderate common effect, the Hispanic Background domain had little in common with the other domains. Most of the effect of the the Children in Child Welfare domain was also unique.

Table 15

Regression Results: Effect of Domains on Original Index Scores

Domain (x)	R	R^2	R^2_{adj}	B	p	*Unique*	*Common*	*Total*
	1.0	1.0	1.0					
FS/ED*				.578	.000	.300	.287	.587
HISPANIC BACKGROUND				.626	.000	.363	.051	.414
CHILD WELFARE				.379	.000	.128	.037	.165

Note: N = 77. *Unique* = *x*'s unique effect. *Common* = ∑ *x*'s common effects. *Total* = *Unique + Common.* *Family Structure/Economic Distress.

Regarding the alternate index, Table 16 shows a similar pattern in terms of one domain (Health) having relatively low total effect. The domain of Families appeared to dominate the alternate index, with a unique effect (.46) that far exceeded the next highest unique effect (.12 for the Health domain) and a total effect of .84, which was considerably greater than the next highest total effect (.29 for the Communities domain). In comparison, the largest total effect among original index domains was .59 for Family Structure/Economic Distress. The domain of Communities had very little unique effect on overall scores, but like the Families domain, had a modest common effect. In contrast, only one domain in the original index had a moderate common effect. Like the Children in Child Welfare domain in the original index, most of the effect of the the Health domain was unique, with almost no effect shared with the other domains.

Table 16

Regression Results: Effect of Domains on Alternate Index Scores

Domain (x)	R	R^2	R^2_{adj}	β	p	*Unique*	*Common*	*Total*
	1.0	1.0	1.0					
FAMILIES				.754	.000	.456	.388	.844
COMMUNITIES				.329	.000	.082	.204	.286
HEALTH				.369	.000	.122	.005	.127

Note: Unique = x's unique effect. *Common* = ∑ x's common effects. *Total = Unique + Common.*

Research Question 3

The third research question asked, *What is the relative effect of individual indicators and domains on outcome rankings? In other words, how stable are index rankings when individual indicators or domains are removed from the index?* To answer this question each index was recalculated several times, each time excluding one indicator or the set of indicators comprising each domain. Scores on each reduced index, also referred to as a scenario, were transformed into percentiles. The difference between the reduced and full percentiles represented an indicator's effect on a county. As percentiles are ranks that convey the percentage of cases below a particular ranking, higher percentiles represented greater risk in both indexes. For example, a county at the 95th percentile was considered at higher risk than 95% of the counties in the dataset.

A decrease in risk ranking with the removal of a particular indicator (or domain) meant it was driving a county's higher ranking. For example, if a county scored at the 80th percentile on the full index but dropped to the 55th percentile when the poverty indicator was removed from the index calculation, then the county likely had a high rate of childhood poverty that was pulling up its average score. On the other hand, an increase in risk ranking with the exclusion of an indicator meant the indicator was lowering a ranking. If, with the exclusion of the poverty indicator, a county moved in the opposite direction from above, going from the 55th to the 80th percentile, then the county likely had a low poverty rate that pulled down the average score on the full index. For this analysis, shifts in rank from the full index of 15 to 19 were considered moderate, while those at approximately 20 or above were considered significant. The terms rank, position and percentile are used interchangeably throughout this discussion.

Overall impact. Two approaches were used in the analysis of indicator and domain impact. The first examined impact overall, and the second examined impacts on individual counties. This section reviews the results of overall impact, with shifts in rank presented in absolute values. The discussion of county-level impact examines direction of rank changes.

Indicators. The median and maximum percentile shifts in rank for each indicator-reduced scenario for the original index are listed in Table 17, along with the number of counties that experienced notable shifts (≥ 15 percentiles).

Table 17

Impact on Ranks with Indicators Excluded One at a Time: Original index

Domain*	Excluded indicator	Median rank shift	Maximum rank shift	Number of counties shifting ≥ 15 ranks
HB	Hispanic	2.6	20.8	3
FS/ED	American Indian	3.9	19.5	3
HB	Migrant	1.3	33.8	2
CW	Foster care	2.6	24.7	2
CW	Abuse/neglect	2.6	22.1	2
FS/ED	Maternal age	3.9	16.9	2
HB	ELL	2.6	36.4	1
FS/ED	Single parent	2.6	18.2	1
FS/ED	Poverty	2.6	18.2	1
HB	Maternal education	2.6	14.3	0

Note: N = 77. *HB = Hispanic Background; FS/ED = Family Structure/Economic Distress; CW = Children in Child Welfare.

With the exception of the migrant indicator, which had a median rank change of 1 percentile when excluded, median shifts in rank were between 3 and 4 percentiles on the reduced scenarios of the original index. This means at least half of the counties moved up or down by no more than four positions across the scenarios. Most indicators significantly impacted one or two counties, with three counties experiencing significant changes in ranks with the exclusion of the American Indian and Hispanic indicators, respectively. Across indicators, the maximum shift in rank ranged from 36 positions with the exclusion of the ELL indicator to 14 positions with the exclusion of the low maternal education indicator.

The median and maximum percentile shifts in rank for each indicator-reduced scenario for the alternate index and the number of counties shifting by 15 or more percentiles are listed in Table 18.

Table 18

Impact on Ranks with Indicators Excluded One at a Time: Alternate Index

Domain	Excluded indicator	Median rank shift	Maximum rank shift	Number of counties shifting ≥ 15 ranks
Communities	ELL	2.6	27.3	7
Health	Prenatal	2.6	22.1	6
Health	Birth weight	3.9	24.7	5
Health	Vaccine	2.6	22.1	5
Families	Maternal education	2.6	29.9	4
Families	Single parent	3.9	27.3	4
Communities	Poverty	3.9	20.8	4
Families	Foster care	2.6	16.9	4
Health	Tobacco	3.9	35.1	3
Families	Abuse/neglect	3.9	20.8	3
Families	Maternal age	2.6	20.8	3

Note: N = 77.

Median shifts in rank again ranged from 3 to 4 percentiles, with similar maximum shifts as the reduced scenarios for the original index. The scenario that excluded tobacco use had the largest maximum shift of 35 positions, while foster care had the smallest maximum shift at 17 positions. From three to seven counties were notably affected by the indicators, more so than across scenarios for the original index. The indexes that excluded tobacco use, young maternal age and abuse/neglect affected the fewest number of counties, while the exclusion of the ELL and prenatal care indicators resulted in the greatest number of counties experiencing moderate to large shifts in rank.

Of the indicators common to both the orignial and alternate indexes, the reduced scenarios for the alternate index resulted in an increase in the number of counties moderately to significantly affected. One additional county was impacted with the exclusion of young maternal age and abuse/neglect, respectively. The alternate scenario that excluded foster care affected two additional counties, followed by three additional counties for the scenarios that excluded low maternal education, single parent and poverty. The ELL indicator experienced the greatest change in impact, with one county meaningfully affected when this indicator was removed from the original index to seven counties shifting by at least 15 percentiles when excluded from the alternate index.

Domains. The median and maximum percentile shifts in rank for each domain-reduced scenario for the original and alternate indexes and the number of counties that experienced moderate to significant shifts are shown in Table 19. Rank changes are again presented in absolute values.

Table 19

Impact on Ranks with Domains Excluded One at a Time: Original and Alternate Indexes

Excluded domain	Median percentile shift	Maximum percentile shift	Number of counties shifting ≥ 15 positions
Original Index			
Hispanic Background	7.8	75.3	18
Family Structure/Economic Distress	10.4	36.4	24
Children in Child Welfare	5.2	24.8	11
Alternate Index			
Ready Families	18.2	71.4	41
Ready Communities	5.2	37.7	10
Ready Services - Health	6.5	44.2	14

Note: N = 77.

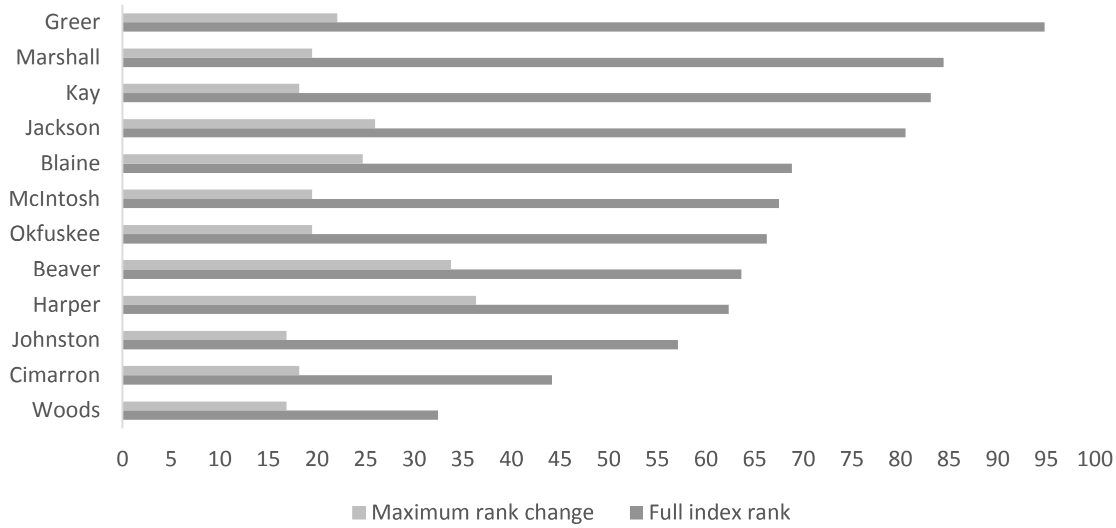
Median shifts across domain-reduced scenarios for the original index ranged from approximately 5 to 10 percentiles, meaning at least half of counties experienced relatively low effects when a particular domain was excluded. In contrast, median shifts in rank across scenarios for the alternate index ranged from 5 to 18 percentiles. Both the Hispanic Background domain in the original index and the Families domain in the alternate index had maximum changes in rank of more than 70 positions. The lowest shift for the original index was 25 positions for the scenario that excluded the Children in Child Welfare domain. For the alternate index, the index that excluded the Communities domain resulted in the lowest shift at 38 positions. The number of counties moving above or below 15 percentiles from their full index rankings was relatively balanced in the original index, with ranges of 11 counties for the Children in Child Welfare domain to 24 counties for the Family Structure/Economic Distress domain. In comparison, more than half of counties (41) shifted by at least 15 positions with the exclusion of the Families domain from the alternate index, while 14 counties were affected by the exclusion of the Health domain and 10 by the exclusion of the Communities domain.

County-level impact. The second part of the analysis of indicator and domain impact focused on individual counties, with the aim of identifying those with the largest shifts in rank on each reduced scenario for indicators and domains. The first part of the analysis for each index uses absolute values, while the second part examines direction of rank changes.

Indicators**.** Figure 1 shows maximum shifts in ranks across the indicator-reduced scenarios, relative to rankings on the original full index, for counties with absolute changes in position of 15 percentiles or more on at least one scenario.

Figure 1

Maximum Rank Change on Indicator-Reduced Scenarios Relative to Full Index Rank: Original Index*



Note: *For counties with shifts in rank of 15 percentiles or more.

Of Oklahoma's 77 counties, 12 experienced moderate to significant shifts, with about a third affected by 2 or 3 indicators. There were 8 counties that experienced changes in rank of approximately 20 or more positions, with 3 experiencing shifts of more than 25 positions. The 2 counties with the largest absolute shifts of approximately 35 percentiles ranked at approximately the 63rd percentile on the full index. Depending on the indicators used in the index and the direction of the effect, the meaning of the rankings for these counties could range from relatively low risk to extremely high risk. In general, higher shifts in ranks on the reduced scenarios generally corresponded to higher rankings on the full index. Of counties with moderate to significant shifts in position, 3 were ranked below the 60th percentile on the full index. Across the scenarios, Alfalfa

County remained the county with the lowest risk ranking, while Harmon County's top ranking was mostly unaffected.

Regarding which indicators drove county rankings and the direction of effects, Table 20 shows counties with shifts in rank of 15 percentiles or more when each indicator was excluded from the original index.

Table 20

Changes in Ranks on Indicator-Reduced Scenarios: Original Index*

Hispanic		ELL		Migrant		Maternal education	
Beaver	-15.6	Harper	-36.4	Jackson	-26.0	none ≥ |15%|	
Marshall	-19.5			Beaver	-33.8		
Jackson	-20.8						
American Indian		**Poverty**		**Single parent**		**Maternal age**	
Cimarron	+16.9					Harper	+16.9
McIntosh	-19.5	Cimarron	-18.2	Kay	-18.2	Johnston	-16.9
Okfuskee	-19.5						
Abuse/neglect		**Foster care**					
Cimarron	+16.9						
Greer	-22.1	Woods	-16.9				
		Blaine	-24.7				

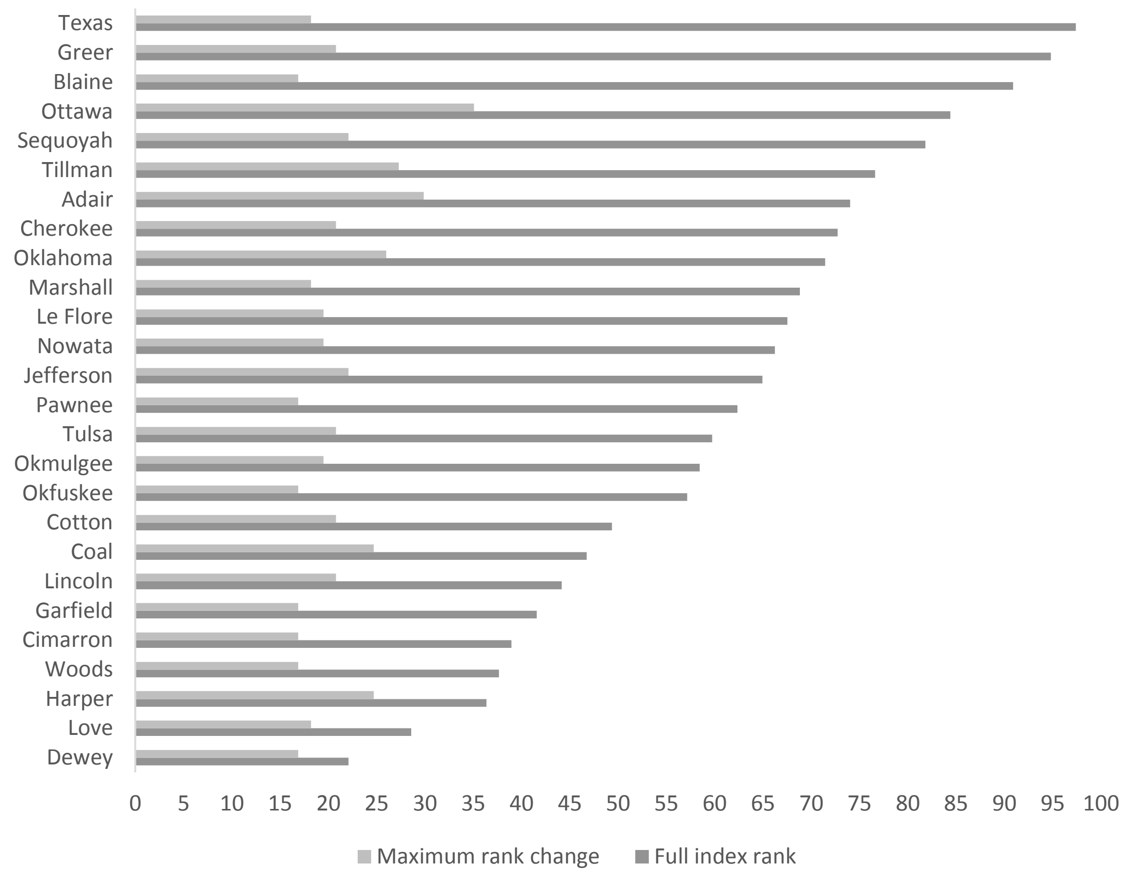
Note: * For counties with shifts in rank of 15 percentiles or more.

For the most part, the exclusion of particular indicators resulted in a decline in risk ranking rather than an increase. Harper County experienced the largest shift, dropping by 36 percentiles with the exclusion of the ELL indicator. This was followed by Beaver and Jackson counties, which declined by 34 and 26 positions, respectively, with the exclusion of the migrant indicator. In addition to Marshall County, both counties also experienced notable declines with the removal of the Hispanic indicator. Blaine County's ranking dropped by 25 positions with the exclusion of the foster care indicator, and McIntosh and Okfuskee both declined in risk ranking by 20 positions with the exclusion of the American Indian indicator. This indicator was one of three that resulted in increases in rank for some counties when excluded, which means these indicators reduced rankings on the full index. For example, with a score at the 44th percentile on the full index, Cimarron County increased in rank to the 61st percentile on the scenarios that left out the American Indian and abuse/neglect indicators. In contrast, Cimarron dropped to the 26th percentile with the exclusion of the poverty indicator.

Figure 2 shows maximum shifts in ranks across the indicator-reduced scenarios, relative to rankings on the full alternate index, for counties with absolute changes in position of 15 percentiles or more on at least one scenario. Twice as many counties experienced notable shifts on the alternate index scenarios compared to those of the original index. Of the 26 counties with moderate to significant shifts in positions, about half were affected by 3 or more indicators. A change in rank of 20 or more positions was experienced by 16 counties, with 4 experiencing shifts of approximately 25 to 35 positions. These counties had ranks on the full index ranging from approximately 70 to 85 percentiles, meaning risk rankings could range from moderately low to extremely high depending on the direction of the effect. The largest shifts across reduced scenarios were again generally associated with higher rankings on the full index, and both Alfalfa and Harmon counties retained their lowest and highest rankings across scenarios. Unlike scenarios associated with the original index, the state's two largest counties, Oklahoma and Tulsa, both experienced significant shifts on at least one of the alternate reduced indexes.

Figure 2

Maximum Rank Change * *on Indicator-Reduced Scenarios Relative to Full Index Rank: Alternate Index*



Note: *For counties with shifts in rank of 15 percentiles or more.

Table 21 shows shifts of at least 15 percentiles above and below full alternate index rankings with the exclusion of particular indicators. Although there were more increases in rank across the reduced scenarios for the alternate index compared to the original index, decreases in rank were again more common. The indexes that excluded poverty and ELL, respectively, had only declines in rankings. All but the index that excluded the foster care indicator had at least one county that increased or decreased in rank by at least 20 positions.

Table 21

Changes in Ranks on Indicator-Reduced Scenarios: Alternate Index*

Maternal age		Maternal education		Abuse/neglect		Foster care	
Tulsa	+20.8	Cotton	+20.8	Cimarron	+15.6	Garfield	+16.9
Oklahoma	+15.6	Okmulgee	+15.6			Tillman	+15.6
Le Flore	-19.5	Marshall	-18.2	Coal	-18.2	Blaine	-16.9
		Adair	-29.9	Greer	-20.8	Woods	-16.9
Single parent		**Poverty**		**ELL**		**Vaccine**	
Lincoln	+20.8					Jefferson	+22.1
Okfuskee	+16.9					Okmulgee	+19.5
Nowata	-15.6	Pawnee	-15.6	Marshall	-15.6	Cotton	-16.9
Tillman	-27.3	Tillman	-15.6	Tulsa	-15.6	Pawnee	-16.9
		Cimarron	-16.9	Tillman	-16.9	Nowata	-19.5
		Cherokee	-20.8	Texas	-18.2		
				Harper	-24.7		
				Oklahoma	-26.0		
				Adair	-27.3		
Birth weight		**Prenatal**		**Tobacco**			
Coal	+24.7	Coal	+18.2	Cimarron	+15.6		
Pawnee	+16.9	Love	+18.2				
Tillman	+15.6						
Jefferson	-16.9	Dewey	-16.9	Jefferson	-19.5		
Nowata	-16.9	Cotton	-18.2	Ottawa	-35.1		
		Tillman	-18.2				
		Sequoyah	-22.1				

Note: *For counties with shifts in rank of 15 percentiles or more.

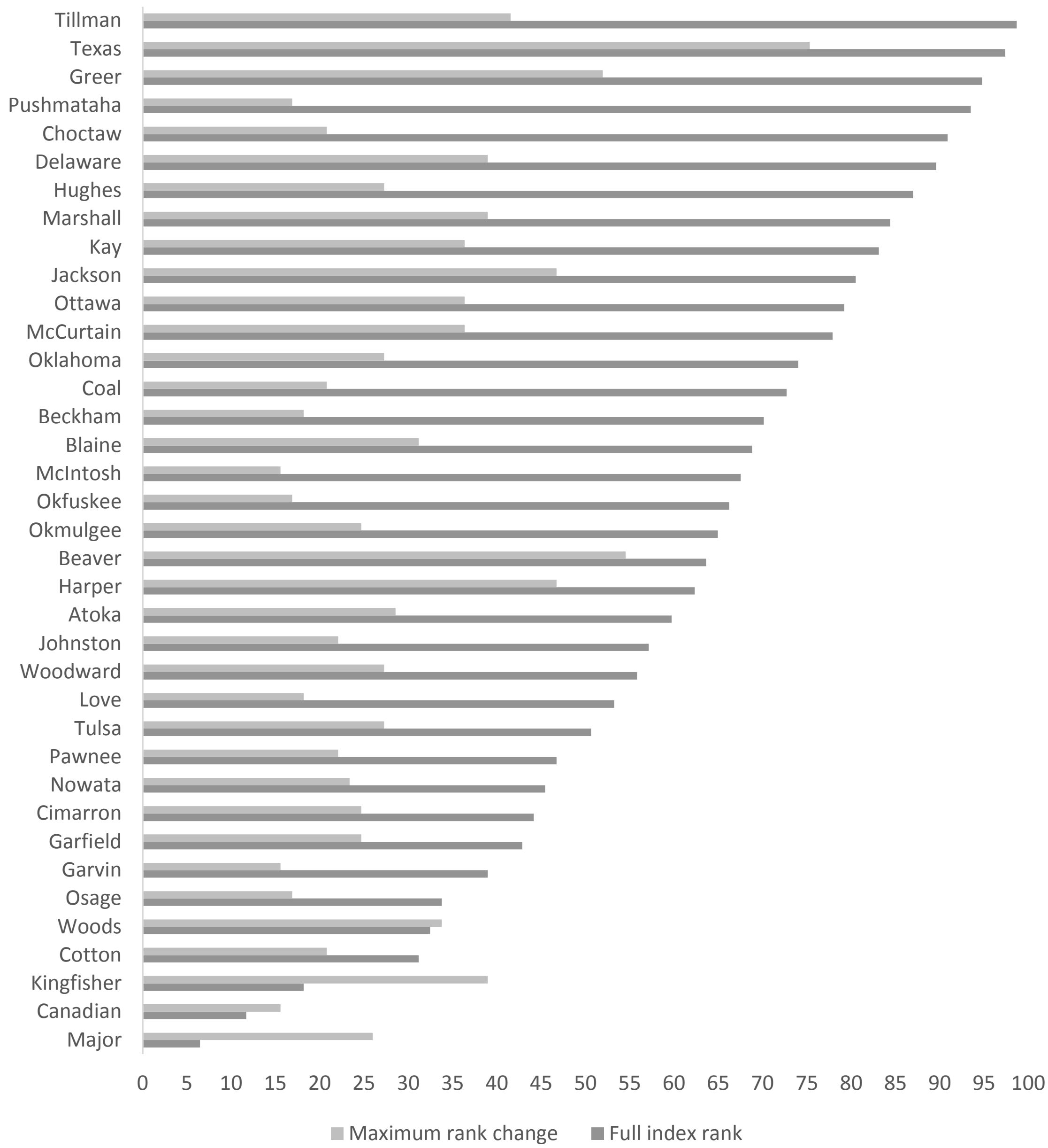
The index that excluded tobacco use had the largest decrease with Ottawa County declining in rank by 35 positions, while the largest increase was for the index that excluded low birth weight, with Coal County moving up in rank by nearly 25 percentiles. Tulsa County had the largest shift on the index that excluded young maternal age, increasing by 21 positions, while Oklahoma County had a moderate increase of 16 positions on this scenario. Conversely, on the ELL-reduced scenario, which saw no increases in rank, Oklahoma County had the second-largest decrease at 26 percentiles, while Tulsa county declined in rank by 16 percentiles. Two other counties (Harper and Adair) had significant declines on the ELL-reduced scenario, with Adair County also largely affected by the low maternal education indicator, declining in rank by 30 positions with its removal.

Seven counties that had notable shifts on the reduced scenarios for the original index also had similar shifts on alternate index scenarios. Four counties (Blaine, Cimarron, Greer and Woods) had ranks on the full original and alternate indexes that were influenced by the abuse/neglect or foster care indicators. For the most part, shifts in position were moderate on both indexes with the exception of Blaine County, which experienced a larger impact from the foster care indicator on the original index. Harper County's risk ranking was significantly affected by the ELL indicator in both indexes, but more so in the original index. Okfuskee County was mostly impacted by the American Indian indicator in the original index, but in the alternate index the single parent indicator had the largest effect. Cimarron County's ranking was largely influenced by the abuse/neglect and poverty indicator on both indexes, and moderately influenced by indicators specific to the original and alternate indexes (American Indian and tobacco use, respectively). While Marshall County was significantly affected by the Hispanic indicator in the original index, it was mostly influenced by the low maternal education indicator in the alternate index.

Domains. Figure 3 shows maximum shifts in ranks across domain-reduced scenarios, relative to rankings on the full original index, for counties with absolute changes in position of 15 percentiles or more on at least one scenario. As expected, counties were more significantly impacted by the exclusion of particular domains than individual indicators. There were 37 counties that experienced shifts of 15 positions or more on at least one reduced scenario, with 29 changing by at least 20 positions. Of these, 23 shifted by 25 percentiles or more. Most counties were moderately to significantly impacted by the exclusion of one domain, while nine counties experienced notable shifts with the exclusion of two domains. Three counties (Beaver, Harper and Woods) were significantly affected by all three domains. As with individual indicators, larger shifts were generally associated with higher rankings on the full index. There were a few exceptions, with three counties ranked near the bottom on the full index experiencing shifts of more than 25 percentiles on at least one reduced scenario. With low rankings, large shifts carry less importance than for counties ranked toward the middle or top. For example, with a ranking on the full index of about 5 percentiles, Major County's shift of 25 positions still places it as relatively low risk. As with the indicator-reduced scenarios, Alfalfa and Harmon County were virtually unchanged from their lowest and highest ranks across scenarios.

Figure 3

Maximum Rank Change on Domain-Reduced Scenarios Relative to Full Index Rank: Original Index*



Note: *For counties with shifts in rank of 15 percentiles or more.

Table 22 shows shifts above and below full original index rankings of at least 15 percentiles with the exclusion of particular domains.

Table 22

Changes in Ranks on Domain-Reduced Scenarios: Original Index*

Hispanic Background		Family Structure/ Economic Distress		Children in Child Welfare	
Woods	+28.6	Kingfisher	+39.0	Harper	+26.0
Cotton	+20.8	Woods	+33.8	Cimarron	+24.7
Nowata	+20.8	Beaver	+29.9	Garfield	+24.7
Okmulgee	+20.8	Harper	+28.6	Beaver	+20.8
Coal	+18.2	Woodward	+27.3	Love	+18.2
Atoka	+16.9	Tulsa	+27.3		
Osage	+16.9	Major	+26.0		
Pawnee	+16.9	Beckham	+18.2		
McIntosh	+15.6	Blaine	+18.2		
		Oklahoma	+18.2		
		Canadian	+15.6		
		Garvin	+15.6		
Kingfisher	-16.9	Pushmataha	-16.9	Beckham	-16.9
Tulsa	-20.8	Choctaw	-20.8	Okfuskee	-16.9
Oklahoma	-27.3	Pawnee	-22.1	Coal	-20.8
Marshall	-39.0	Johnston	-22.1	Woods	-24.7
Tillman	-41.6	Nowata	-23.4	Blaine	-31.2
Harper	-46.8	Okmulgee	-24.7	Greer	-52.0
Jackson	-46.8	Hughes	-27.3		
Beaver	-54.6	Atoka	-28.6		
Texas	-75.3	McCurtain	-36.4		
		Ottawa	-36.4		
		Kay	-36.4		
		Delaware	-39.0		

Note: * For counties with shifts in rank of 15 percentiles or more.

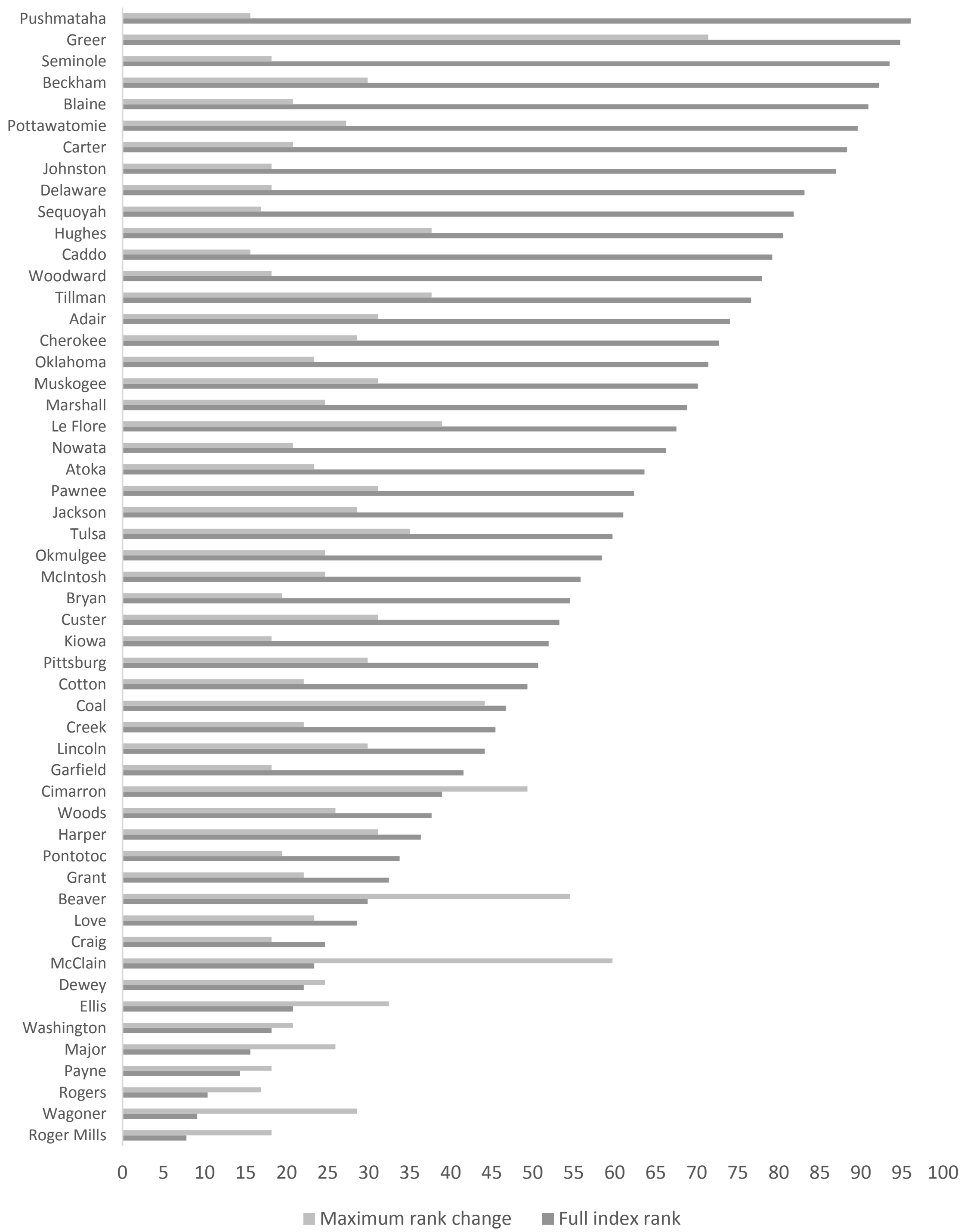
There was a relatively even split in terms of number of counties moving up or down in rank for each scenario. The magnitude of shifts in direction were relatively similar for the removal of the Family Structure/Economic Distress domain, with a range of +/- 39 positions. In comparison, the exclusion of the Hispanic Background domain resulted in several extremely large declines in position, with 6 counties dropping in risk ranking by 39 to 75 percentiles. While Texas county scored at the 97th percentile on the full original index, the removal of the Hispanic Background domain dropped its ranking to the 22^{nd} percentile. In comparison, Texas County's ranking dropped by less than 3 positions with the exclusion of the the Family Structure/Economic Distress and Children in Child Welfare domains. The removal of the latter domain also resulted in more extreme declines in ranking compared to increases, with Blaine and Greer counties dropping by 31 and 52 percentiles, respectively, while the largest increase was 26 percentiles for Harper County.

While Tulsa and Oklahoma counties had no significant shifts on indicator-reduced scenarios for the original index, both experienced significant shifts on the domains. Tulsa County, with a full index score at the 51^{st} percentile, was largely affected by the Hispanic Background domain, dropping to the 30^{th} percentile with its removal, as well as the Family Structure/Economic Distress domain, increasing to the 78^{th} percentile with its removal. Oklahoma County, with a full index score at the 74^{th} percentile, was less affected by the latter domain but dropped to the 47^{th} percentile with the exclusion of the Hispanic Background domain. Alfala and Harmon counties, the lowest and highest ranked on the full index, were again mostly unaffected by the exclusion of the domains.

Figure 4 shows maximum shifts in ranks across domain-reduced scenarios, relative to rankings on the full alternate index, for counties moving at least 15 positions.

Figure 4

Maximum Rank Change on Domain-Reduced Scenarios Relative to Full Index Rank: Alternate Index*



Note: *For counties with shifts in rank of 15 percentiles or more.

There were again considerably more significant impacts of excluding domains from the alternate index than excluding indicators. The removal of at least one domain impacted two-thirds of counties, for a total of 53 moderately to significantly affected. There were 40 counties that changed rank by approximately 20 or more positions; of these, 28 shifted by at least 25 percentiles. The majority of counties were moderately to significantly affected by one domain, while eight counties were impacted by two domains. Two counties (Blaine and Ellis) experienced large changes on all three domains. While the county with the largest shift ranked near the top on the full index, in general there was no clear pattern of shifts in relation to full index ranks. Of the counties that changed ranks by more than 25 percentiles, 6 had shifts that were larger than their ranks on the full index, which ranged from approximately 10 to 40 percentiles. As with the original index, for many counties the exclusion of a particular domain drastically changed the meaning of their risk ranking. The lowest and highest rankings of Alfalfa and Harmon counties, respectively, continued to be mostly unaffected by the exclusion of any domains.

Table 23 shows shifts above and below full index rankings of at least 15 percentiles with the exclusion of particular domains. As shown in earlier analyses of domain impacts, the alternate index is heavily influenced by the Families domain, which is comprised of the maternal education, maternal age, single parent, abuse/neglect and foster care indicators. This is demonstrated by the large number of counties with moderate to significant shifts relative to the other two domains. There were 10 counties notably affected by the Communities domain and 14 by the Health domain.

Table 23

Changes in Ranks on Domain-Reduced Scenarios: Alternate Index*

Ready Families†				Ready Communities		Ready Services - Health	
McClain	+59.7	Pushmataha	-15.6	Creek	+22.1	Coal	+44.2
Beaver	+54.6	Johnston	-18.2	Cotton	+22.1	Harper	+31.2
Cimarron	+49.4	Delaware	-18.2	Lincoln	+16.9	Love	+23.4
Tulsa	+35.1	Adair	-18.2	Woods	+16.9	Washington	+20.8
Ellis	+32.5	Craig	-18.2			Pontotoc	+19.5
Pawnee	+31.2	Seminole	-18.2			Custer	+19.5
Lincoln	+29.9	Bryan	-19.5			Adair	+15.6
Wagoner	+28.6	Pontotoc	-19.5				
Jackson	+28.6	Carter	-20.8	Cimarron	-16.9	Caddo	-15.6
Major	+26.0	Atoka	-23.4	Oklahoma	-23.4	Jackson	-16.9
Dewey	+24.7	Marshall	-24.7	Cherokee	-28.6	Sequoyah	-16.9
Harper	+22.1	Okmulgee	-24.7	Harper	-29.9	Woodward	-18.2
Grant	+22.1	McIntosh	-24.7	Adair	-31.2	Nowata	-20.8
Garfield	+18.2	Woods	-26.0	Tillman	-37.7	Blaine	-20.8
Roger Mills	+18.2	Pittsburg	-29.9			Pottawatomie	-27.3
Creek	+18.2	Beckham	-29.9				
Kiowa	+18.2	Muskogee	-31.2				
Payne	+18.2	Custer	-31.2				
Rogers	+16.9	Hughes	-37.7				
		Le Flore	-39.0				
		Coal	-41.6				
		Greer	-71.4				

Note: *For counties with shifts in rank of 15 percentiles or more. †Because of the number of counties affected by the Families domain, increases and decreases are presented side-by-side.

Rank changes with the exclusion of the Families domain ranged from a decline of 71 percentiles for Greer County to an increase of 60 percentiles for McClain County. Of the 41 counties notably affected by this domain, 12 had rank changes of less than 20 positions, while 15 changed by approximately 30 or more positions. In contrast, shifts when the Communities domain was excluded ranged from a decrease of 38 percentiles for Tillman County to an increase of 22 percentiles for Creek County. For the Health domain, shifts ranged from a drop of 27 positions for Pottawatomie County to an increase of 44 positions for Coal County. Whereas Texas County had an extreme effect from the

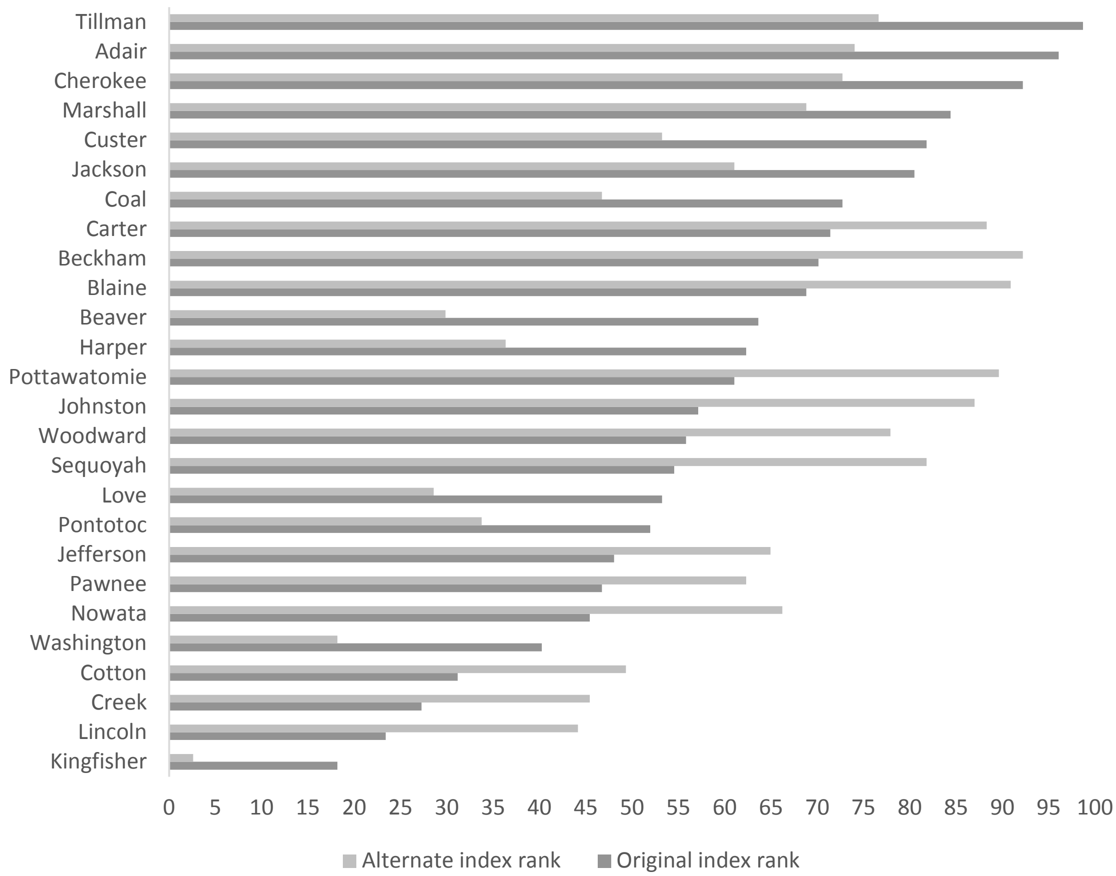
Hispanic Background domain in the original index, its largest change in rank was a decline of 10 positions with the exclusion of the Communities domain, which included the ELL and poverty indicators. Tulsa and Oklahoma counties each had large shifts on one scenario, with Oklahoma declining from the 74^{th} to the 48^{th} percentile with the exclusion of the Communities domain, and Tulsa experiencing an extreme increase from the 60^{th} to the 95^{th} percentile with the exclusion of the Families domain.

Full index rank comparisons. To conclude the analysis of the effect of indicators on outcome rankings, the association between the full original and alternate indexes was assessed. Shifts in county rankings between the two indexes were compared for the extent to which they resulted in changes to county risk group classifications. The four risk groups used to classify counties on the original index were high, high-medium, medium-low and low risk.

The Spearman rank correlation between the two indexes was strong and positive ($r_s = 0.88$), with a median absolute change in rank of 8 percentiles. Figure 5 shows percentile ranks on the original and alternate indexes, ordered by original index rankings. There were 26 counties that changed rank by at least 15 percentiles, with an even split between the number that increased and decreased in rank. Of these counties, 8 shifted by at least 25 percentiles, with 5 decreasing in rank from the original to the alternate index while 3 increased. The largest increase was for Johnston County at 30 percentiles, while Beaver County experienced the largest decrease at 34 percentiles. With a few exceptions, increases in rank on the alternate index occurred for counties with relatively moderate to low rankings on the original index, while the opposite was true for counties that decreased in rank. There was relatively little change from the original to the alternate index for the state's four largest counties. Tulsa and Comanche counties increased in risk ranking by approximately 9 and 7 percentiles, respectively, while Cleveland and Oklahoma counties were virtually unchanged.

Figure 5

Comparison of Ranks on Original and Alternate Indexes for Counties with the Largest Shifts*



Note: *For counties with shifts in rank of 15 percentiles or more.

Table 24 lists changes in rank for counties that increased or decreased from the original to the alternate index by 15 or more percentiles. With the exception of four counties, rank changes were large enough to move counties up or down one risk group. Tillman County retained its high risk grouping on the alternate index despite a decline of 22 positions on the alternate index. Both Creek and Cotton counties remained in the moderately low risk group with an increase of 18 positions, and Kingfisher County continued to be classified as low risk with a decline of 16 positions on the alternate index. Six counties with shifts in rank from approximately 14 percentiles down to extremely minor shifts of 3 percentiles also experienced changes in risk groupings.

Table 24

Changes in Ranks from Original to Alternate Index*

County	Rank increase	Quartile change	County	Rank decrease	Quartile change
Johnston	+29.9	+1	Beaver	-33.8	-1
Pottawatomie	+28.6	+1	Custer	-28.6	-1
Sequoyah	+27.3	+1	Coal	-26.0	-1
Beckham	+22.1	+1	Harper	-26.0	-1
Blaine	+22.1	+1	Love	-24.7	-1
Woodward	+22.1	+1	Tillman	-22.1	0
Nowata	+20.8	+1	Washington	-22.1	-1
Lincoln	+20.8	+1	Adair	-22.1	-1
Creek	+18.2	0	Jackson	-19.5	-1
Cotton	+18.2	0	Cherokee	-19.5	-1
Jefferson	+16.9	+1	Pontotoc	-18.2	-1
Carter	+16.9	+1	Marshall	-15.6	-1
Pawnee	+15.6	+1	Kingfisher	-15.6	0

Note: * For counties with shifts in rank of 15 percentiles or more.

Research Question 4

The last question addressed by this study asked, *To what extent do changes to the indicators and domains affect associations of index rankings with a proxy outcome of school readiness risk?* To answer this question, Spearman rank correlations were computed between percentiles for each full and reduced index and for entering kindergartners scoring below proficiency on state-mandated pre-literacy assessments.

Table 25

Spearman Correlation Coefficients for Ranks on Literacy Nonproficiency and Full and Reduced Indexes

	Literacy		**Literacy**
Original index	**.20**	**Alternate index**	**.21**
Excluded indicator		*Excluded indicator*	
Maternal age	.18	Maternal age	.20
Maternal education	.22	Maternal education	.21
Abuse/neglect	.18	Abuse/neglect	.19
Foster care	.18	Foster care	.21
Single parent	.17	Single parent	.15
Poverty	.20	Poverty	.18
ELL	.22	ELL	.20
Hispanic	.24	Vaccinations	.24
American Indian	.19	Birth weight	.22
Migrant	.21	Prenatal care	.20
		Tobacco use	.22
Excluded domain		*Excluded domain*	
Hispanic Background	.26	Ready Families	.01
Family/Economics*	.08	Ready Communities	.17
Children in Child Welfare	.14	Ready Services - Health	.24

Note: N = 77. *Family Structure/Economic Distress.

As shown in Table 25, both the original and alternate indexes were modestly correlated with literacy nonproficiency (r_s = .20 and r_s = .21, respectively), as were the various scenarios. For both indexes, the lowest correlations with indicator-reduced scenarios were for those that excluded the single parent indicator (r_s = .17 and r_s = .15, respectively). Regarding domains, the exclusion of the Family Structure/Economic Distress domain from the original index resulted in an extremely low correlation (r_s = .08). This domain included three of the four indicators that resulted in the lowest correlations when excluded from the original index. Removing the Ready Families domain from the alternate index resulted in virtually no relationship with the proxy outcome variable (r_s = .01). Of the five indicators included in this domain, two resulted in the lowest correlations with literacy nonproficiency when excluded from the alternate index one at a time. For the original index, the highest correlation with indicator-reduced scenarios was with the index that excluded the Hispanic indicator (r_s = .24), while the removal of the Hispanic Background domain resulted in the highest correlation (r_s = .26) among domain-reduced scenarios. For the alternate index, the highest correlation with indicator-reduced scenarios was with the index that excluded the vaccinations indicator (r_s = .24). The exclusion of the Health domain resulted in the same coefficient, which was the highest among domain-reduced scenarios. The exclusion of the abuse/neglect and foster care indicators from the original index resulted in slightly lower correlations with the proxy outcome than the full index, while on the alternate index only the abuse/neglect indicator lowered the correlation coefficient.

Chapter Summary

This chapter presented the results of analyses related to the four research questions and concludes with a summary of findings. The implications of these findings are discussed in Chapter 5. No significance tests were conducted as the dataset represented the population of all counties in Oklahoma.

With the exception of two indicators that were deemed to be effects of starting school unprepared rather than predictors of risk, the same indicators considered for inclusion in the original index were also considered for the alternate index. Indicators were examined for outliers and descriptive statistics, and zero-order bivariate correlations were reviewed for the extent to which they supported including particular indicators. The original index contained more indicators that had multiple outlier counties and had more moderate to high correlations than the alternate index (see Table 7). On the other hand, the alternate index had more negative correlations than the original index. Negative correlations between indicators are acceptable if both indicators are meaningful to the construct (Diamantopoulos & Winklhofer, 2001). The final indicators included in the alternate index were grouped conceptually into three domains, while the indicators included in the original index were grouped into three domains on the basis of statistical analysis.

Correlational and commonality analysis of indicators and domains with overall index scores yielded similar information. No indicators appeared to clearly dominate either index by themselves. Across both the original and alternate indexes, the indicators of young maternal age, low maternal education, poverty and single parent were the most important, having the highest correlations with and largest effects on overall scores. While both the maternal age and maternal education indicators were highly important to both indexes, they traded places to a small extent in terms of impact, with maternal education having the largest effect on the original index and young maternal age having the largest effect on the alternate index.

The abuse/neglect and foster care indicators increased in importance in the alternate index, while the impact of the ELL indicator declined. In the original index, all three indicators had the same moderate associations and had basically the same small effect on overall score. Conversely, in the alternate index, the abuse/neglect and foster care indicators increased moderately in terms of correlations with overall score, while the association of ELL with overall score declined considerably. In the alternate index, the indicators of birth weight and prenatal care were more highly correlated with overall score than was ELL. While the relationships of domains and index scores were fairly balanced in the original index, the Ready Families domain clearly dominated the alternate index. This is unsurprising given this domain includes three of the four most important indicators noted above. Both indexes had one domain that had little to no effect on overall scores relative to the other domains.

The impact of indicators and domains on risk rankings showed that the alternate index had more counties notably affected by changes to the indicator set than the original index. In both indexes, a few changed rank by at least 25 positions. Regarding the seven indicators that were common to both indexes, there was an increase in the number of counties that experienced meaningful changes in rank when these indicators were excluded from alternate index score calculations. Four of these indicators, however, had lower maximum rank changes on the alternate index. While about half of all counties experienced notable changes in rank with the exclusion of domains from the original index, nearly two-third had meaningful shifts when a domain was removed from the alternate index. Both indexes had one domain causing shifts as extreme as 75 positions when excluded. As expected, removing the Families domain had at least a moderate impact on over half the state's counties.

The original and alternate indexes were highly correlated, with the majority of counties experiencing relatively low shifts in rank from one to the other. However, of the state's 77 counties, nearly one-third experienced notable shifts, with 10% changing ranks significantly between indexes. Of the 23 counties that were outliers on at least one indicator in either the original or alternate index, 10 experienced shifts in rank from 15 to 34 positions, with all but one ranking lower on the alternate index compared to the original. Of the nine outlier counties that experienced a notable drop in rank, eight were outliers on at least one of the migrant, Hispanic and ELL indicators, with four being outliers on two or all three indicators. The only outlier county to have a significant increase in rank on the alternate index was an outlier on the prenatal indicator.

The analysis of indicator impact concluded by examining the associations of full and reduced indexes to a proxy outcome of being unprepared for school. Both the original and alternate indexes had modest correlations with ranks (in percentiles) of entering kindergartners scoring below proficiency on pre-literacy assessments. Across reduced scenarios for both indexes, there were small changes in correlations, with the largest decreases associated with the removal of the single parent indicator as well as its associated domains, while the largest increases were associated with the removal of the vaccinations and Hispanic indicators, as well as their associated domains.

CHAPTER V

CONCLUSION

"I have been struck again and again by how important measurement is
to improving the human condition."
-- Bill Gates (2013)

Starting school unready to learn has serious consequences for a child's future. A considerable body of research has demonstrated that children who start behind often stay behind. Opportunities for a full and productive life can become limited before a child even steps through the doors of a kindergarten classroom. Several states, including Oklahoma, have developed indexes to identify areas where children are at greatest risk for starting school without the developmental foundation necessary for academic success. These indexes are intended for use by policy makers and early childhood stakeholders in efforts to mitigate factors that impede development and address inequities in access to quality early education and child care. As such, school readiness risk indexes serve as powerful information tools for improving quality of life for children and their families by supporting evidence-based decision making. Given the importance of such indexes and the myriad possible indicators that could be used in their construction, an assessment of the impact of indicator selection on index rankings was warranted.

The purpose of this study was to examine the sensitivity of the Oklahoma School Readiness Risk Index (OK SRRI) to changes to the indicator set used in its construction and to assess the extent to which various sets affected associations with a proxy measure of starting school unready to learn. All analyses were conducted twice, once with the original index and again with an alternate index. Initial indicators considered for the original index came from a review of the empirical literature on school readiness risk factors, with indicators selected for analysis on the basis of data availability. This same set was also considered for the alternate index. For the original index, the final set of indicators and the domains used to group them were selected on the basis of principal components analysis and multiple linear regression. The final set for the alternate index was selected based on analysis of descriptive statistics, with domains formed in context of a theoretical understanding of child development as the product of interrelated factors, the accumulation of which magnifies risk for starting school unprepared.

Of the 10 indicators used in the original index and the 11 selected for the alternate, 7 were common to both indexes. These indicators related to maternal characteristics, economics, education and child maltreatment. The primary difference was that the original index contained four indicators related to race/ethnicity and none related to health, while the alternate index included one indicator related to race/ethnicity and four related to health. Further, the domain structures of both indexes differed, with the original index containing a Hispanic Background domain and the alternate including a Health domain. Both had domains related to family characteristics. The alternate index used the same methods of scale transformation (*z*-scores), weighting (equal) and aggregation (arithmetic mean of all scores) as the original index.

This study was guided by the following research questions:

1. What is the relationship among scores on the overall index, the domains and individual indicators?
2. What is the impact of the indicators and domains on the overall index? In other words, do one or more indicators or domains dominate the index?
3. What is the relative effect of individual indicators and domains on outcome rankings? In other words, how stable are index rankings when individual indicators or domains are removed from the index?
4. To what extent do changes to the indicators and domains affect associations of index rankings with a proxy outcome of school readiness risk?

Summary of Findings

One of the greatest challenges to index construction is that discussion of the actual process of selecting indicators is limited, and there is little guidance regarding under what circumstances an indicator should be excluded (Diamantopoulos & Winklhofer, 2001). While procedures have been recommended for analyzing the underlying impact of indicators on index rankings (Nardo et al., 2008; Saisana, 2008; Saisana & Munda, 2008), there is ambiguity on what to do when indicators are found to have a large impact. The simple solution of excluding them is problematic given that indexes follow a formative rather than reflective measurement model, meaning indicators define, rather than reflect, the construct being measured, making these indicators causal in nature. Thus, if an indicator is found to dominate an index, yet it is an important part of the theoretical framework of the phenomenon under consideration, removing it would change the construct. While increasing attention has been paid to this issue in regard to index construction, there seems to be a gap between those who advance methods for assessing indicator impact (Nardo et al., 2008; Saisana, 2008) and those who promote the concept of indexes as defining a particular construct (Bollen & Lennox, 1991; Diamantopoulos & Winklhofer, 2001).

Given the above limitation, this study did not aim to produce an absolute conclusion regarding which indicators should be used in Oklahoma's risk index. As stressed by Nardo et al. (2008, p. 23), "there may be no single definitive set of indicators" for any index. Rather, the purpose of this study was to contribute to the transparency of the OK SRRI by conducting an analysis of its sensitivity to changes to the set of indicators as well as domains. This section provides a summary of the main findings, which are descriptive rather than empirical, and discusses reasons for differential impacts of indicators when used in various combinations. Because this study used the population of counties in Oklahoma, no statistical significance testing was conducted.

Synopsis of Major Findings

Although there was a strong association between the original and alternate indexes, with a Spearman rank order correlation of $r_s = .88$, approximately one-third of the state's counties shifted rank by at least 15 percentiles between the indexes. This cutpoint was used to examine the magnitude of shifts in rankings between indexes as well as across various reduced scenarios that excluded one indicator or domain at a time from score calculations. Shifts between 15 and 19 percentiles were considered moderate, while shifts of 20 percentiles or above were considered significant. Eight counties experienced shifts in rankings between the original and alternate indexes of at least 25 percentiles. For both indexes, higher percentile rankings represented greater risk. Compared to the original index, more counties had moderate to significant shifts in rank on the reduced scenarios of the alternate index. The same was true of the exclusion of sets of indicators that comprised particular domains. For most counties with at least moderate shifts between the two full indexes, risk groupings increased or decreased by one level, with the majority of changes occurring between the medium-high and high risk groups. A few counties with shifts as small as 3 percentiles also experienced changes in risk groupings, while other counties with large shifts remained in the same group.

Regarding associations with a proxy outcome of school readiness risk, entering kindergartners behind on literacy skills, both indexes had similar correlations of about $r_S =$ .20, with associations across reduced scenarios increasing or decreasing to a relatively small degree. The exceptions were the reduced scenarios for both the original and alternate indexes that excluded the family-related domains, which resulted in virtually no association with literacy skills. Each of these domains included three of the following four indicators – low maternal education, young maternal age, poverty and single parent – in different combinations. Among indicator-reduced scenarios, the exclusion of the single parent indicator resulted in the largest drop in correlation for both indexes.

Regarding overall associations of indicators and domains with index scores and the extent to which any single indicator or domain dominated an index, similar results were found for the two methods used to assess these characteristics. First, scores for individual indicators and domains were correlated with overall index scores. This was followed by multiple regression commonality analysis to determine the amount of variance in overall index scores uniquely explained by an indicator or domain and the amount shared with other indicators or domains. Of indicators common to both the original and alternate indexes, the same four indicators noted above related to maternal education and age, poverty and family structure emerged as the most important in terms of their associations with and effects on overall scores for both indexes. The indicators of abuse/neglect and foster care increased in magnitude from the original to the alternate index, while the ELL indicator declined considerably in terms of its association with and total effect on index scores. The smallest correlations between individual indicators and original index scores were for the American Indian and migrant indicators ($r < .32$), while the alternate index had small correlations with ELL, tobacco use and vaccinations ($r < .20$). Of the Health domain indicators in the alternate index, prenatal care had the highest correlation with overall score ($r = .36$), followed by low birth weight. All of the above indicators had small effects on their respective full indexes. The vaccinations indicator was negatively correlated with alternate index score, and, in addition to tobacco use, was negatively associated with other indicators in terms of shared effects. Negative commonality coefficients result from suppression or negative associations between independent variables used in a regression model, and indicate that one variable confounds the effects of other variables (Pedhazur, 1997, Nimon, 2010).

Regarding the three domains in the original index, none appeared to dominate, while of the three domains in the alternate index, the Families domain clearly dominated with a near perfect correlation with overall score and an extremely large effect relative to the other two domains. For both indexes, the correlations among domains were mostly smaller than correlations between domains and overall scores. This is expected and desirable as domains should account for particular dimensions of a multidimensional construct, yet there should be some degree of overlap. Both indexes, however, had one negative correlation between domains, which is undesirable as it suggests domains work against each other. For both indexes, there was at least one domain that had a small effect relative to the other domains. In the original index, the two domains (Hispanic Background and Children in Child Welfare) had primarily unique effects, while the same was true in the alternate index for one domain (Health).

Understanding differential impact of indicators. Examining the impact of indicators common to both indexes at the county level showed these indicators had different effects when used with a different set of indicators. The differential performance between indexes can be partly explained by the presence of multiple high correlations between indicators and the presence of more outliers in the original index compared to the alternate. The problem with highly correlated indicators is that they are assumed to tap the same or similar dimension, which makes them duplicative; therefore, they contribute to double counting a single aspect of a construct when equal weights are used (Nardo et al., 2008). Moreover, outliers can make a considerable impact on index rankings, especially when associated with highly correlated indicators. The Hispanic and ELL indicators, with a correlation of r = .80, had 7 and 10 outliers, respectively, while the migrant indicator, which was moderately correlated with Hispanic, had 6. In comparison, abuse/neglect and foster care, with basically the same correlation coefficient as between Hispanic and ELL, had relatively few outliers and considerably less variability than the latter indicators.

Two counties, Tillman and Greer, serve as case studies into the impacts of high correlations, outliers and variability. Tillman County was ranked at the 99th percentile on the original index, compared to the 77th on the alternate index. The county did not experience any shifts across reduced scenarios on the original index of more than 15 percentiles, but had such shifts on six reduced scenarios of the alternate index. The most significant shift stemmed from the impact of the single parent indicator, which, when excluded, dropped Tillman's ranking by 27 percentiles on the alternate index. In comparison, the exclusion of the single parent indicator from the original index resulted in a decline of only 1 percentile. The difference in indicator performance across the indexes was due to the varying magnitude of Tillman's z-scores on the raw variables. A z-score indicates how many standard deviations an observation is above or below the mean. Tillman had a mean score on the original index of $z = 1.22$ and a range of scores across the index of $z = -1.6$ for abuse/neglect to $z = 6.7$ for migrant, far above the next highest score. Despite being an extreme outlier and therefore inflating the mean, removing the migrant indicator from the calculation of Tillman's index score resulted in a drop of only 11 positions. Across remaining indicator-reduced scenarios, the county's ranking was mostly unchanged. This is explained by the high scores on four other indicators: single parent ($z = 2.4$), Hispanic ($z = 2.2$) and ELL and poverty ($z = 1.8$), while remaining scores were moderate to low.

When the Hispanic Background domain, comprised of Hispanic ethnicity, ELL, migrant and low maternal education, was removed from Tillman's score calculation on the original index, it dropped by 42 positions. In contrast, when the Children in Child Welfare and/or Family Structure/Economic Distress domains were excluded, Tillman's ranking was essentially unchanged. Although the latter domain contained two of the county's highest *z*-scores, single parent and poverty, its remaining two indicators had low or negative scores. Therefore, it is clear that the migrant indicator, as well as the strong relationship between the Hispanic and ELL indicators, made a considerable impact on Tillman's overall ranking relative to other indicators despite a very small rate of migrant children. This example shows why it is important that indicators have sufficient coverage across the population of counties in order to be included in an index. Only six counties had any reported rates of migrant children, which made all six outliers.

In the alternate index, which excluded the Hispanic and migrant indicators, Tillman's *z*-scores ranged from z = -2.8 for birth weight to z = 2.4 for single parent for a mean score of .25. Although still ranked in the high risk group, Tillman's shifts in rankings across reduced scenarios were more illuminating of the county's major issues. In addition to the large impact of the single parent indicator, the county also had moderate declines in ranking when poverty and ELL were excluded, pointing to Tillman's relatively high rates on these indicators. In addition, Tillman had a relatively large drop with the exclusion of the prenatal care indicator, demonstrating that late or no prenatal care is an important issue for this county.

Greer County serves as another example of the impact of high correlations, outliers and variability. Greer had the 10th highest rate of Hispanic children (26%) but the 54th highest rate of ELL children (1%), thus the relatively high correlation of these indicators was not an issue for this county. In addition to having the highest rate of births to teen mothers (22%), Greer also had the highest rates of confirmed cases of abuse/neglect and children placed in foster care at approximately 8% for each. This county was the only one to be an outlier on both of the latter indicators, with scores of z = 5.1 and z = 3.6, respectively. In comparison, its scores for the Hispanic and young maternal age indicators were z = 1.4 and z = 2.5, respectively, and z = -.5 for ELL. Greer's average score on the original index was z = .76 (95th percentile). In comparison, removing either of the abuse/neglect or foster care indicators but retaining the other dropped Greer's ranking to z = .44 and z = .27, respectively. Moreover, Greer's ranking dropped by 52 positions with the exclusion of the Child Welfare domain from the original index. Despite Greer's high rates of births to teen mothers, the combined effect of abuse/neglect and foster care clearly drove its overall ranking on both the full and alternate indexes. These indicators tap the same dimension, child maltreatment, which means this dimension implicitly carried twice the weight in the overall score.

The difference in magnitude of z-scores relative to the proportions of children for the above indicators also exemplifies the impact of variability on county rankings. With limited variability, high or low rates can result in extreme z-scores relative to indicators with greater variability. The standard deviations and range of observations for the Hispanic and ELL indicators were about seven times as large as for the child welfare indicators. Therefore, a rate of 8% on one indicator can result in a considerably larger score than rates of 25% on another. The reason the impact of both indicators increased in general on the alternate index while the ELL indicator declined is because of the issues described above. The effect of the ELL indicator was magnified by its correlation with the Hispanic indicator in the original index, while excluding this indicator from the alternate index implicitly increased the weight of the child maltreatment indicators.

Therefore, as Mayer (2008, p. 287) contends, “the simple choice of inclusion affects the weighting of data, which can have a large impact on the resulting value of the index.”

Of the top five ranked counties in the original and alternate indexes, three were common to both indexes. Of the top four counties in the original index, all had the highest rates of ELL children, with the three highest ranked also having the highest rates of Hispanic children and representing two of the six counties with migrant children. In contrast, while Harmon and Texas counties held onto the highest and third highest overall rankings on the alternate index, two of the top five counties had among the lowest rates for the Hispanic, ELL and migrant indicators. Both counties, however, had among the highest rates of births to teen mothers, single parents and children in foster care, while one county also ranked near the top in terms of poverty and abuse/neglect rates. Greer was the fifth highest ranked county in both indexes.

Just as counties with high rates on the Hispanic- or child welfare-related indicators were “punished” with high risk rankings, counties with low rates on these indicators were just as likely to be “rewarded” with a lower overall risk ranking, despite the possibility that they may have high rates on indicators not included in the index. For example, Johnston County increased from the 57th percentile on the original index to the 87th percentile on the alternate index. Its largest *z*-score was for young maternal age (z = 1.9), followed by low birth weight (z = 1.4), while its scores on the Hispanic, ELL, migrant, American Indian and abuse/neglect indicators were at or below the mean. The addition of the low birth weight indicator increased its overall score enough to move it from the medium-high to the high risk group. Every county with moderate to significant increases in rank on the alternate index compared to the original followed a similar pattern with at least one relatively high score on the health indicators, which were not highly associated. In contrast, several counties that had moderate or significant declines in rank on the alternate index, such as Tillman, were strongly influenced by the multiple use of racial/ethnic-related indicators in the original index. Due to variability issues discussed above, there was considerably less impact for the child welfare indicators. For counties with moderate to large shifts in rank between the indexes, the meaning of their risk rankings changed considerably with the use of a different indicator set.

Discussion of Findings

A brief discussion of what these findings mean for indicator selection is in order. Although there were a few small associations between indicators, as well as domains, and overall index scores, small associations do not necessarily suggest an indicator should be removed from an index. An issue known to be related to a construct and problematic across a particular area may still warrant inclusion. For example, limited English proficiency is a major stumbling block to success in school. The Oklahoma State Department of Education (OSDE) reports that approximately 10% of Oklahoma’s pre-kindergarten and kindergarten children are in ELL programs, which makes this indicator important to monitor even if it only contributes a small amount to overall index scores. Moreover, according to the Oklahoma State Department of Health (OSDH) and the Centers for Disease Control (CDC), the state has close to 10% of infants born without appropriate prenatal care, higher than the nation, and just under that born at low birth weight. Because of these issues, it is important that indicators not be excluded without consultation with experts in the field (Saisana, 2008).

The fact that the vaccinations indicator was both negatively associated with index score and confounded the effects of other indicators, however, is an indication that this variable may not work well despite its association with child health and therefore school readiness. Current issues surrounding vaccinations and parental decisions to not vaccinate their children may help explain this. Among Oklahoma counties, vaccination rates in some of the poorest counties are higher than counties with lower poverty rates. This may be due in part to these families having access to vaccines through Medicaid or the federally funded Vaccines for Children program for low-income families (Cosgrove, 2015, February 12). Lower rates among more affluent counties may be partially explained by Oklahoma's law allowing parents to request an exemption to required vaccinations for personal or religious reasons ("Oklahoma Immunization Act," 1970). Regarding the reliability of indicator data, not all clinics in Oklahoma are required to report to the state's immunization registry, thus the data likely have some degree of bias.

Findings related to the association of domains with overall scores suggest that domains may need to be re-examined and indicators re-grouped to assess how changes to domain sets alter their associations with index scores. For example, the Health domain may be improved by replacing the vaccinations indicator with others available at the county level. Data related to food insecurity and uninsured children are available from the KIDS COUNT data center, while data on preterm births are available from the OSDH online data query system. The abuse/neglect and foster care indicators, used in the Ready

Families domain, could be moved to the Health domain to account for emotional and physical aspects of health and possibly reduce the dominance of the Families domain. These indicators comprise their own domain in the original index.

The effects of the Hispanic Background domain are difficult to interpret given its inclusion of the low maternal education indicator. Although approximately one-third of the nation's children whose mothers lack a high school diploma are Hispanic (Child Trends, 2014), this issue is certainly not exclusive to the Hispanic community. With three of the four indicators in the Hispanic Background domain related to Hispanic ethnicity, and thus carrying three-fourths of the weight in this domain, the maternal education indicator likely had a small impact on overall domain scores.

Other child well-being and school readiness risk indexes have grouped maternal education with family-related domains. For example, the KIDS COUNT Index (Mather & Dupuis, 2012) includes a domain labeled Family and Community, comprised of indicators related to undereducated parents, births to teens, single parent families and children in high-poverty areas. Pennsylvania's school readiness risk index places maternal education with teen birth rate and children born to young and single mothers in the Maternal Risk domain, while Louisiana's index organizes maternal education with teen birth rate and poverty in the Economic Risk domain. Moving the poverty indicator to the Families domain in the alternate index would leave the ELL indicator by itself. This suggests the need for additional indicators to create a new domain, which could relate to education. The KIDS COUNT Index, based on the premise that indicators specific to very young children as well as those that indicate future outcomes are important to child well-being, includes such a domain. Considering indicators in the context of a theoretical framework that encompasses all dimensions of a particular phenomenon is imperative for creating a balanced representation of a multidimensional construct. Further, Mather and Dupuis (2012, p. 3) note that domains should be structured in a way that makes them "relevant for state-level policy."

The fact that some counties with minor shifts in rank between the original and alternate indexes changed risk groupings while a few with significant shifts did not suggests that the process for risk groupings may need to be re-examined. Groupings should be able to capture meaningful shifts in rank, and more variability in groupings means that counties on the edge of categories see smaller changes in the interpretation of their risk ranking. For example, the use of quintiles over quartiles would put about 15 of the state's 77 counties in each group, corresponding approximately to the cutpoint of 15 percentiles used in this study to denote meaningful shifts. The Child Opportunity Index (COI) used quintiles to categorize communities in terms of very low, low, moderate, high and very high opportunity (Acevedo-Garcia et al., 2014).

Finally, with the exception of the ELL indicator, the alternate index excluded indicators related to race and ethnicity. This is not meant to imply that race and ethnicity should be excluded from the analysis of school readiness risk, but rather they should take on different form, such as examining access by racial/ethnic groups to quality early education and child care. As demonstrated by the COI, rankings can be mapped onto neighborhoods to better understand the situations of children across diverse racial and ethnic groups. Of six indexes, other than the OK SRRI, related to child well-being and school readiness reviewed for this study, only one included an indicator specific to race/ethnicity, measured by percent non-white children in a county.

Implications

This study demonstrated that changes to the set of indicators used in an index can change the meaning of a construct, which has the potential to considerably alter the interpretation of a particular ranking. This underscores the significance of the processes used to select indicators. The findings from this study support the results of sensitivity analyses of other indexes that found at least some significant shifts in rank with the use of different indicator sets (Chakraborty et al., 2005; Houweling et al., 2003; Jones & Andrey, 2007; Mather & Dupuis, 2012). This raises an important issue related to the literature on index construction. One method of examining indicator dominance is to exclude indicators and examine changes to rankings. It is recommended that these reduced indexes should have little change in rankings (Saisana, 2008). An argument could be made, rather, that this merely suggests indicators are interchangeable, and therefore may not be measuring distinct dimensions of a multidimensional construct. As demonstrated by this and other index-related studies, indicators do not move in relation to each other and not all counties are alike. It seems inevitable that rankings for some counties will be largely affected by the set of indicators used, while others will remain the same regardless. Therefore, it is expected that an index, if it includes indicators that tap distinct aspects of a construct, will be at least somewhat sensitive to changes to the indicator set.

A review of the literature on index construction shows that many recommendations regarding index construction and validation efforts fail to consider how indexes relate to reflective and formative measurement models. Many approaches to index development and validation follow a reflective model, for which the core assumption is that indicators reflect, or are the effect of, a particular unidimensional construct. Effect indicators, therefore, are considered to be interchangeable, and those selected to model a latent (unobserved) variable represent a sample of all possible indicators that reflect the construct. As such, indicators are assumed to be highly correlated. Indexes developed and purported to be validated using these assumptions risk misrepresenting a particular situation. Given the political nature of indexes, this can have significant consequences in terms of inadequately distributed resources and erroneous policy decisions.

The choice of indicators used to define a social phenomenon can direct resources toward or away from particular issues. For example, excluding health indicators from an index related to child well-being means that health-related issues will be ignored when it comes to distributing resources or setting policy on the basis of index rankings. Of the four state indexes of school readiness risk examined for this study, as well as the national KIDS COUNT Index, Oklahoma's is the only index that lacks any indicators directly related to health. The KIDS COUNT Index measures child well-being across states, with higher rankings representing a better situation for children. Of the index's four domains of Economic Well-Being, Education, Health, and Family and Community Context, Oklahoma's worst ranking was on the Health domain, with only five states ranked lower on this domain.

Indicator selection can also drive resources toward or away from particular counties. This has considerable implications for several counties in Oklahoma, as the primary purpose of its risk index is to compare risk levels with the reach of early education and child care programs. The intention is to identify counties where risk is high but availability of quality programs is low. If resources were to be allocated based on this information, several counties would be treated differently depending on whether the original or alternate index was used. For example, Adair County ranked as a high risk county in the original index, but dropped to the medium-high risk group in the alternate index. Conversely, Sequoyah ranked as medium-high risk in the original index, but increased to the high risk group in the alternate index. Based on six indicators of child care access and quality, these counties were in the lower reach groups (Lazarte Alcalá & Schumacher, 2014). If high risk counties of low to moderate child care reach were selected for resource allocation, then Sequoyah, with high rates of infants born to mothers with inadequate prenatal care, would be excluded from consideration with the use of the original index. In contrast, Adair, with the highest rate of American Indian children in the state, would be excluded from consideration using the alternate index.

The issue of where to direct resources is also relevant to decisions regarding how to group counties according to risk, as well as the need for a participatory approach to index development. It may be that smaller groupings that create more variability in terms of number of groups would be more informative than larger groups, especially if resources are tight. Moreover, whether either of the indicators noted above would be agreed upon by stakeholders as being part of the construct of school readiness risk in Oklahoma is not known.

While Oklahoma’s index aims to inform decisions related to early care and education, such an index can also be useful for addressing the causes of being at risk for starting school unprepared. One issue that must be kept in mind when considering index rankings is that a high overall ranking does not imply high rates on all indicators. Understanding indicator impact can provide information regarding which risk factors are more problematic for one area over another, which informs resource allocation. For example, if a county has similarly high rates on ELL and limited prenatal care, and it is ranked at extremely high risk, how do the resources get distributed in a climate of limited funds? If larger shifts in rankings occur with the exclusion of the ELL indicator, then, relative to other counties, ELL is a more pressing issue than prenatal care, while prenatal care might emerge as the more critical issue for another county. Therefore, sensitivity to changes to the indicator set may not necessarily be a mark against the quality of an index. An important caveat, though, is that the degree to which indicator sensitivity meaningfully informs decision making depends on the set of indicators used.

In proposing an index to measure disaster preparedness, Simpson (2006, p. 5) argues that “the act of deciding what to count is value oriented and subjective in nature.” This is particularly salient with a construct such as school readiness risk that lacks a universally accepted definition in terms of what, exactly, constitutes risk. What constitutes risk in one community may be different from what constitutes risk in another. Instead of relying on statistical analysis alone, indicators should follow a theoretical framework, and decisions regarding which indicators to keep and which to exclude should be informed by those with a stake in the outcomes (Simpson, 2006). For example, involving leaders in early childhood programs as well as representatives of programs designed to mitigate school readiness risk factors can result in an index that has community buy-in.

Several indexes related to various social issues, including child well-being, have used a participatory approach to selecting indicators that involved community members and experts in the field (Diamantopoulos & Winklhofer, 2001; Freebairn & King, 2003; Mather & Dupuis, 2012; Roy, Chan, & Rainis, 2014; Thivierge et al., 2014). These approaches have taken several forms, including surveys, focus groups, interviews and Q-methodology (Doody, Kearney, Barry, Moles, & O'Regan, 2009), a method that combines qualitative and quantitative approaches to studying individual viewpoints. Boyd and Charles (2006) set forth a framework for a community-level indicator development process that includes creating a common vision, establishing a theoretical framework to reflect the vision, and identifying relevant characteristics of the proposed social issue.

Limitations

An overarching limitation in regard to indicator selection is data availability at the desired spatial level and timeframe, particularly if indexes are to be used to track trends over time. This study used only the set of indicators examined for the original index as well as the same data, although there are other indicators related to school readiness risk with data available at the county level and more recent data available for the indicators that were used. There is also one indicator, tobacco use during pregnancy, no longer publicly available at the county level. Given that Oklahoma was the only state among 40 where rates of pre-pregnancy smoking increased from 2000 to 2010 (Tong et al., 2013), this is unfortunate and highlights the problems of limited data availability.

Restricting the study to a particular set of indicators limited the organization of domains in the alternate index, which followed, to the extent possible, the domain structure of NSRII's recommended school readiness indicators. These indicators did not include ELL, however, so this indicator was somewhat arbitrarily grouped with poverty under the Ready Communities domain with the idea that community readiness means being ready to provide English-language instruction to children who need it. However, this makes for a somewhat obscure domain. As domains are useful for pinpointing areas of concern and promoting appropriate responses, there should be no ambiguity to their interpretation (Mather & Dupuis, 2012). Data are further limited by reliability issues. A few indicators used administrative data, which Rouse and Fantuzzo (2009, p. 11) maintain lack the "scientific checks and balances to ensure the information is reliable and valid." Data for other indicators came from the American Community Survey of the U.S. Census Bureau, which has long acknowledged issues with data collection that result in undercounts, particularly among young children (Tate, 2012; U.S. Census Bureau, 2014).

This study was further limited in that it used the same methods of normalization, weighting and aggregation applied to the original index, limiting the examination of sensitivity to only one aspect of index construction. While the equal weighting scheme is the most transparent and is widely used, it is also likely to introduce bias and, as demonstrated earlier, can strongly influence final rankings (Mayer, 2008). Moreover, when the arithmetic mean is used as the overall index score, negatively correlated indicators can result in indicator scores essentially cancelling each other out. This is particularly problematic when indicators are expected to be at least somewhat associated and relates to the problem of ecological fallacy.

An issue for this study as well as the field of index construction in general, although rarely addressed (Schmidtlein et al., 2008), ecological fallacy occurs when aggregated (ecological) data are used to make inferences regarding individuals. Although the explicit purpose of a school readiness risk index is not to make inferences about individual children, to a certain extent this is exactly what it does. It is not counties themselves that are of interest, but rather the children who live in those counties. To say that children in a particular county are at greater risk for starting school unprepared than children in another county assumes that all children in each county have the same level of risk. This is an individual-level inference and the crux of the ecological fallacy problem because indicators influential at the individual level may not have the same magnitude or direction of relationship at the population level. For example, there was a nearly perfect positive correlation between the raw number of infants born at low birth weight and the number born to mothers with late or no prenatal care. However, the correlation between the percentages was moderately small and negative. Ecological correlations rely on the sum (marginal frequency) of individuals with particular characteristics for categories, e.g., late or no prenatal care and low birth weight, as well as the sum of all individuals in those categories (e.g., all live births). The sums could result from any number of sets of internal frequencies. For example, if 50% of infants were born to mothers with late or no prenatal care, and 50% of infants were born at low birth weight, then at the population level there is a perfect association between these indicators. There is no reason to assume, however that such a one-to-one association exists at the individual level. Infants with limited prenatal care may or may not be the same infants born at low birth weight. In other words, marginal frequencies do not “fix” the individual data points used for an individual-level correlation, meaning a single correlation could have associated with it several different individual-level correlations (Robinson, 2009, p. 339).

As this research was essentially a case study, using data for one state to examine sensitivity to changes to the indicator set, specific results regarding indicator performance cannot be generalized to other indexes of school readiness risk. Moreover, as the findings demonstrate, indicators perform differently when used in conjunction with different indicator sets, so no absolute judgments can be made regarding the “best” set of indicators. Related to this is the decision to set a cutpoint for meaningful shifts in rank across reduced scenarios of 15 percentiles. There is some consensus in the literature that shifts of about 20 percentiles or more are at least moderately significant, but there are no hard and fast rules from which to make judgements. The cutpoint used for this study was selected to provide for a more in-depth examination of indicator impact on county rankings, but the degree to which shifts of 15 percentiles are meaningful is up for debate. Moreover, with quartiles used to group counties into risk categories, there were counties that shifted by 15 percentiles or more that failed to move into another risk category. Therefore, the extent to which shifts in rankings affected risk groupings is in part an artifact of the method used to create groups.

Finally, validation is a prominent limitation of indexes, as classic validation efforts apply to the reflective measurement model used to develop scales as opposed to the formative model that underlies index construction. Recommendations to overcome this obstacle include linking rankings with a latent variable theorized to be a consequence or antecedent of the index score, or composite variable (Bollen & Bauldry, 2011; Diamantopoulos & Winklhofer, 2001). The latent variable should consist of several effect indicators believed to reflect the composite variable and hence are expected to be correlated. Ideally, for a school readiness risk index, this would consist of several measures of development, from cognitive to social-emotional domains. As there is only one state-mandated assessment for entering kindergartners, this study was restricted to linking index scores with only one indicator, the percent of kindergartners scoring below proficient on a literacy assessment. As discussed in Chapter 1, literacy is only one aspect of what makes a child prepared for school. A firm conclusion regarding associations between this outcome and reduced and full indexes is further limited by problems with interpreting effect sizes (e.g., correlation coefficients) as the extent to which a judgment can be made about their magnitude is debatable. Cohen (1988) provided conventions for interpreting effect sizes. A correlation coefficient of .10 is considered small, .30 is medium and .50 as large. However, Cohen stressed that these are general guidelines and the true meaning of an effect size varies depending on the topic and design of a study.

Directions for Future Research

There is considerable room for further research in the field of index construction, particularly concerning indexes related to school readiness risk and child well-being in general. As there are other possible indicators that could be used in Oklahoma's index, an examination of the impact of various other combinations of indicators and domains could greatly inform efforts to select an optimal (as opposed to a definitive) set for the state. Before additional indicators are considered, however, analysis of the effects of using a participatory approach to indicator selection on the index is warranted. Such an approach would ideally involve community leaders, policy makers, educators, program directors, parents, clinicians and other stakeholders and could inform a variety of questions. For example, does a participatory approach result in a particular conceptualization of school readiness? If so, do different indicators emerge that were not initially considered in constructing the index, and does this change the domain structure? Another question could consider the impacts of using different types of participatory approaches. For example, the use of surveys or focus groups could be compared to Q methodology, provided enough indicators were available at the county level and regularly updated. Questions that could be answered through such a study include the extent to which different approaches: 1) changed participants' understanding of school readiness risk; 2) facilitated a consensus regarding a definition of risk for the state of Oklahoma; and 3) increased awareness and use of the index by policy makers and other stakeholders.

A participatory approach could also be used to investigate which factors stakeholders perceive to be the most salient for their communities, thus identifying how risk manifests in particular areas. These perceptions could be compared with county risk rankings, using a particular set of indicators, to determine the extent to which there is congruence with community perceptions. In light of limitations on validation efforts, local knowledge is important for ensuring the most appropriate indicators are used (Schmidtlein et al., 2008; Simpson, 2006). A participatory approach could also be used to examine the extent to which results are presented in a format that is useful to policy makers and others.

Another avenue for future research includes validation efforts using approaches recommended specifically for indexes, such as linking outcomes to construct antecedents and/or consequences. A latent variable comprised of multiple reflective indicators could be modeled, given availability of data, to represent being unprepared at the start of school. Reflective indicators could also be used to estimate a multiple indicators multiple causes (MIMIC) model that would account for the interrelationships among indicators used to model such a variable and those used to construct the index.

As several other states have developed indexes to measure school readiness risk, the present study could be replicated with these indexes. The sensitivity of Oklahoma's index to different methodological approaches, such as the use of different weights or aggregation methods, would also contribute to a more thorough understanding of how choices made during different stages of index construction affect index rankings. This would also inform the extent to which different methodological approaches could overcome some of the problems related to highly correlated indicators or those with limited variability as well as outliers. Another direction for research is to carry out commonality analyses on domains in an effort to determine whether domains themselves are dominated by particular indicators and the extent to which different domain structures result in differential impact of indicators on scores. This is particularly relevant if scores on domains, rather than individual indicators, were used to derive overall index scores. Even when an equal weighting scheme is used for all indicators, an implicit weighting occurs when different numbers of indicators are used in domains. Therefore, understanding which indicators contribute the most to domain scores and which contribute the least can inform understanding of how domains impact overall index scores. Moreover, as domains should represent distinct aspects of a construct, knowing whether some indicators confound the effects of other indicators in a single domain can be useful for determining the extent to which domains follow a multidimensional framework as theorized.

Concluding Remarks

Indexes have the potential to be powerful tools in raising awareness of important societal problems and to inform policy and resource decisions in response to these issues. As such, the processes of defining and measuring a problem are crucial if an index is to be widely accepted as a trusted source of information. However, unlike the field of scale development, there is little agreement on a "best" method of index development. This limitation is compounded by the fact that much of the guidance on assessing index validity and reliability derives from psychometric theory, which is used to assess the extent to which scales tap latent psychological traits. Social issues, however, are not psychological constructs. The same methods used to measure a psychological construct and ensure confidence in the outcomes are not applicable to the measurement of a social issue. One ramification of this is that rather than sampling a group of indicators from a domain of all possible indicators, as is the process for scale development, indexes require a census of indicators that cover the entire scope of an issue (Bollen & Lennox, 1991). As error cannot be accounted for in indexes as it can in scales, excluding some indicators that comprise part of the construct introduces a large degree of error, the ramifications of which cannot be clearly known. Grace and Bollen (2008, p. 201) contend that if an index represents only part of its associated construct, then it must be “recognize[d] that it is an imperfect measure of that construct.”

The final set of indicators selected, and how they are selected, matters for community buy-in. Several studies on the indicator selection process have found that in addition to resulting in a holistic collection of indicators valuable for decision making, “the process of engaging people to select indicators also provides an opportunity for community empowerment that conventional development approaches have failed to provide” (Fraser, Dougill, Mabee, Reed, & McAlpine, 2006, p. 114). Acceptance of an index as a trusted source of information cannot occur in the absence of negotiation among policy makers and other stakeholders. (Gaye, 2007).

The only way to truly assess the extent to which a school readiness risk index relates to school readiness is through the use of individual-level data on children that can be matched between schools and social service agencies. For example, the Kids Integrated Data System (KIDS) in Philadelphia matches data on individual children across the Philadelphia school district with the city’s human services, health, and housing agencies (Fantuzzo & Perlman, 2007). Unfortunately, such a system is not available in Oklahoma or in many other states. Although the Oklahoma Department of Human Services received funding to investigate the validity of the index as a measure of risk using individual-level data, accessing data has proven problematic, with matching data across education and social services even more challenging. Until such studies are possible, Oklahoma’s index is the only means of identifying where children are at greatest risk. Therefore, it is critical that there be general agreement on the definition of risk and its various dimensions. Statistical techniques of indicator selection are inherently flawed if the indicators examined fail to represent a social construct in its totality. Without a strong theoretical foundation, indexes run the risk of being purely "exercises in measurement" rather than meaningful contributors to understanding a particular social construct (Freudenberg, 2003, p. 29).

The findings from this study make a significant contribution to the field of index construction, particularly in terms of measuring risk for poor school readiness, by demonstrating that changes to the indicator set change the construct. This study supported findings from other research that "type of risk matters," as different risk factors have different impacts (Rouse & Fantuzzo, 2009). In light of the sensitivity of the index to changes to the indicator set, the Oklahoma index should proceed with caution and be coupled with stakeholder and expert guidance to ensure representations of school readiness risk are reasonable and consistent with local knowledge. Collaborations with organizations dedicated to child well-being, such as the Oklahoma Institute for Child Advocacy, the state grantee for the KIDS COUNT data center, and Smart Start Oklahoma, dedicated to the issue of school readiness, could greatly inform continued index development. There are a considerable number of other organizations and stakeholders that should be consulted as well.

As a sensitivity analysis, this study contributed to increasing the transparency of Oklahoma's risk index. Of all index considerations, transparency is perhaps the most important and most demanded by policy makers (Nardo et al., 2005; Saisana, 2008;

Sharpe & Andrews, 2012). Therefore, understanding the strengths and weaknesses of an index, as well as its biases, is critical to ensuring policy-related decisions rest on a solid foundation (Mayer, 2008). In discussing the primary failures of sustainability indexes, Mayer (2008, p. 288) notes that first and foremost, indexes fail when there is "a lack of consensus on what sustainability is in a quantitative sense." This leads to the three overarching conclusions to this study. First, the choice of indicators has a considerable impact on the exact construct being measured. Second, the process of identifying and selecting indicators should include those with a stake in the outcomes. Finally, in the process of deciding which indicators to ultimately use, "indicator elimination – by whatever means – should not be divorced from conceptual considerations" (Diamantopoulos & Winklhofer, 2001, p. 273).

References

Abrahamse, A. F., Morrison, P. A., & Waite, L. J. (1988). *Beyond stereotypes: Who becomes a single teenage mother?* Santa Monica, CA: RAND Corporation.

Acevedo-Garcia, D., McArdle, N., Hardy, E. F., Crisan, U. I., Romano, B., Norris, D., . . . Reece, J. (2014). The child opportunity index: improving collaboration between community development and public health. *Health Affairs, 33*(11), 1948-1957.

Amato, P. R. (2001). Children of divorce in the 1990s: An update of the Amato and Keith (1991) meta-analysis. *Journal of Family Psychology, 15*(3), 355-370.

Appleyard, K., Egeland, B., van Dulmen, M. H. M., & Alan Sroufe, L. (2005). When more is not better: The role of cumulative risk in child behavior outcomes. *Journal of Child Psychology and Psychiatry, 46*(3), 235-245.

Ards, S. D., Myers Jr, S. L., Malkis, A. E., & Zhou, L. (2003). Racial disproportionality in reported and substained child abuse and nelgect: An examination of systematic bias. *Children and Youth Services Review, 25*(5–6), 375-392.

Bagozzi, R. P. (1994). Structural equation models in marketing research: Basic principles. In R. Bagozzi (Ed.), *Principles of marketing research* (pp. 317-385). Oxford: Blackwell.

Barnett, J., Lambert, S., & Fry, I. (2008). The hazards of indicators: Insights from the Environmental Vulnerability Index. *Annals of the Association of American Geographers*, *98*(1), 102-119.

Barnett, W. S., & Yarosz, D. J. (2007) Who goes to preschool and why does it matter? *Preschool Policy Matters, Issue 8*. New Brunswick, NJ: National Institute for Early Education Research.

Barth, R. P., Wildfire, J., & Green, R. L. (2006). Placement Into foster care and the interplay of urbanicity, child behavior problems, and poverty. *American Journal of Orthopsychiatry, 76*(3), 358-366.

Bartlett, J. D., & Easterbrooks, M. A. (2012). Links between physical abuse in childhood and child neglect among adolescent mothers. *Children and Youth Services Review, 34*(11), 2164-2169.

Bauman, L. J., Silver, E. J., & Stein, R. E. K. (2006). Cumulative social disadvantage and child health. *Pediatrics, 117*(4), 1321-1328.

Behrman, J. R., & Rosenzweig, M. R. (2004). Returns to birthweight. *Review of Economics and Statistics, 86*(2), 586-601.

Bloch, M., Tong, V. T., Novotny, T. E., England, L. J., Dietz, P. M., Kim, S. Y., . . . Tolosa, J. E. (2010). Tobacco use and secondhand smoke exposure among pregnant women in low- and middle-income countries: a call to action. *Acta Obstetricia et Gynecologica Scandinavica, 89*(4), 418-422.

Bloom, D. E., Canning, D., & Weston, M. (2005). The value of vaccination. *World Economics, 6*(3), 15-39.

Bollen, K. A., & Bauldry, S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological Methods, 16*(3), 265-284.

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*(2), 305-314.

Borden, K. A., Schmidtlein, M. C., Emrich, C. T., Piegorsch, W. W., & Cutter, S. L. (2007). Vulnerability of U.S. cities to environmental hazards. *Journal of Homeland Security and Emergency Management*, *4*(2).

Boyd, H., & Charles, A. (2006). Creating community-based indicators to monitor sustainability of local fisheries. *Ocean & Coastal Management, 49*(5), 237-258.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual review of psychology*, *53*(1), 371-399.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (5th ed., pp. 993-1029). New York: Wiley.

Bronfenbrenner, U., & Morris, P. A. (2007). The bioecological model of human development *Handbook of Child Psychology*: John Wiley & Sons, Inc.

Brooks-Gunn, J., Guo, G., & Furstenberg Jr., F. F. (1993). Who drops out of and who continues beyond high school? A 20-year follow-up of black urban youth. *Journal of Research on Adolescence*, *3*(3), 271-294.

Burchinal, M., Roberts, J. E., Zeisel, S. A., Hennon, E. A., & Hooper, S. (2006). Social risk and protective child, parenting, and child care factors in early elementary school years. *Parenting*, *6*(1), 79-113.

Burchinal, M. R., Peisner-Feinberg, E. S., Pianta, R., & Howes, C. (2002). Development of academic skills from preschool through second grade: Family and classroom predictors of developmental trajectories. *Journal of School Psychology, 40*(5), 415-436.

Burchinal, M. R., Roberts, J. E., Hooper, S., & Zeisel, S. A. (2000). Cumulative risk and early cognitive development: A comparison of statistical risk models. *Development Psychology, 36*(6), 793-807.

Burchinal, M. R., Roberts, J. E., Zeisel, S. A., & Rowley, S. J. (2008). Social risk and protective factors for African American children's academic achievement and adjustment during the transition to middle school. *Developmental Psychology, 44*(1), 286-292.

Camargo, K., & Grant, R. (2015). Public health, science, and policy debate: Being right is not enough. *American Journal of Public Health, 105*(2), 232-235.

Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., & Pan, Y. (2014). Early childhood investments substantially boost adult health. *Science, 343*(6178), 1478-1485.

Carlton, M. P., & Winsler, A. (1999). School readiness: The need for a paradigm shift. *School Psychology Review*, *28*(3), 338.

Carneiro, P., & Heckman, J. (2003). Human capital policy, IZA Discussion paper series, No. 821.

Carter, V. B. (2010). Factors predicting placement of urban American Indian/Alaskan Natives into out-of-home care. *Children and Youth Services Review, 32*(5), 657-663.

Chakraborty, J., Tobin, G. A., & Montz, B. E. (2005). Population evacuation: assessing spatial variability in geophysical risk and social vulnerability to natural hazards. *Natural Hazards Review, 6*(1), 23-33.

Chandra, P. C., Schiavello, H. J., Ravi, B., Weinstein, A. G., & Hook, F. B. (2002). Pregnancy outcomes in urban teenagers. *International Journal of Gynecology & Obstetrics, 79*(2), 117-122.

Chen, X.-K., Wen, S. W., Fleming, N., Demissie, K., Rhoads, G. G., & Walker, M. (2007). Teenage pregnancy and adverse birth outcomes: A large population based retrospective cohort study. *International Journal of Epidemiology, 36*(2), 368-373.

Child Trends. (2014). *Indicators on children and youth: Parental education*. Retrieved from http://www.childtrends.org/?indicators=parental-education

Claessens, A., Duncan, G., & Engel, M. (2009). Kindergarten skills and fifth-grade achievement: Evidence from the ECLS-K. *Economics of Education Review, 28*(4), 415-427.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Coley, R. L. (1998). Children's socialization experiences and functioning in single-mother households: The importance of fathers and other men. *Child Development, 69*(1), 219-230.

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research, 61*(12), 1250-1262.

Comfrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Lawrence Erlbaum.

Comparison of Literacy First and DIBELS assessments. (2005). Retrieved May 27, 2014, from http://www.catapultlearning.com/wp-content/uploads/2013/06/LFtoDibels_Comparison.pdf

Cosgrove, J. (2015, February 12). More Oklahoma parents opt out of vaccinating their children, *The Oklahoman*. Retrieved from http://www.oklahoman.com/article/5392432?access=b2c9afd158a602aebb0ac9548dd73fd0

Courtney, M., & Skyles, A. (2003). Racial disproportionality in the child welfare system. *Children and Youth Services Review*, *25*(5–6), 355-358.

Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social vulnerability to environmental hazards. *Social Science Quarterly*, *84*(2), 242-261.

Cutter, S. L., Burton, C. G., & Emrich, C. T. (2010). Disaster resilience indicators for benchmarking baseline conditions. *Journal of Homeland Security and Emergency Management*, *7*(1).

Davidson, R. A., & Lambert, K. B. (2001). Comparing the hurrican disaster risk of U.S. coastal counties. *Natural Hazards Review*.

Deater-Deckard, K., Dodge, K. A., Bates, J. E., & Pettit, G. S. (1998). Multiple risk factors in the development of externalizing behavior problems: Group and individual differences. *Development and Psychopathology*, *10*(03), 469-493.

Desai, M., Fukuda-Parr, S., Johansson, C., & Sagasti, F. (2002). Measuring the technology achievement of nations and the capacity to participate in the network age. *Journal of Human Development, 3*(1), 95-122.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage

Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A Comparison and empirical illustration. *British Journal of Management, 17*(4), 263-282.

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: an alternative to scale development. *Journal of Marketing Research, 38*(2), 269-277.

Diener, E., & Suh, E. (1997). Measuring quality of life: Economic, social, and subjective indicators. *Social Indicators Research, 40*(1-2), 189-216.

Doody, D. G., Kearney, P., Barry, J., Moles, R., & O'Regan, B. (2009). Evaluation of the Q-method as a method of public participation in the selection of sustainable development indicators. *Ecological Indicators, 9*(6), 1129-1137.

Ducan, G. J., & Magnuson, K. A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *The Future of Children, 15*(1), 35-54.

Early Childhood Data Collaborative. (2014). *2013 State of States' Early Childhood Data Systems*. Retrieved from http://www.ecedata.org/files/2013%20State%20of%20States'%20Early%20Childhood%20Data%20Systems.pdf

Eide, E. R., Showalter, M. H., & Goldhaber, D. D. (2010). The relation between children's health and academic achievement. *Children and Youth Services Review, 32*(2), 231-238.

Eisinga, R., Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health, 58*(4), 637-642.

Erickson, M., & Egeland, B. (1996). Child neglect. In J. Briere, L. Berliner, J. Bulkley, C. Jenny & T. Reid (Eds.), *The APSAC Handbook on Child Maltreatment* (pp. 4-20). Thousand Oaks, CA: Sage.

Esty, D. C., Levy, M., Srebotnjak, T., & de Sherbinin, A. (2005). *2005 Environmental Sustainability Index: Benchmarking national environmental stewardship*. New Haven: Yale Center for Environmental Law and Policy. Retrieved from http://www.yale.edu/esi/ESI2005_Main_Report.pdf

Evans, G. W. (2004). The environment of childhood poverty. *American Psychologist, 59*(2), 77-92.

Fantuzzo, J., & Perlman, S. (2007). The unique impact of out-of-home placement and the mediating effects of child maltreatment and homelessness on early school success. *Children and Youth Services Review, 29*(7), 941-960.

Fantuzzo, J. W., LeBoeuf, W. A., & Rouse, H. L. (2014). An investigation of the relations between school concentrations of student risk factors and student educational well-being. *Educational Researcher, 43*(1), 25-36.

Fekete, A. (2009). Validation of a social vulnerability index in context to river-floods in Germany. *Natural Hazards and Earth System Sciences, 9*, 393-403.

Felner, R. D., & DeVries, M. L. (2013). Poverty in childhood and adolscence: A transactional-ecological approach to understanding and ehnacing resilience in contexts of disadvantage and developmental risk *Handbook of Resilience in Children*.

Felner, R. D., Felner, T. Y., & Silverman, M. M. (2000). Prevention in mental health and social intervention. In J. Rappaport & E. Seidman (Eds.), *Handbook of community psychology* (pp. 9-42). New York, NY: Springer.

Finch, B. K. (2003). Socioeconomic gradients and low birth-weight: Empricial and policy considerations. *Health Services Research, 38*(6), 1819-1841.

Findley, S. E., Irigoyen, M., & Schulman, A. (1999). Children on the move and vaccination coverage in a low-income, urban Latino population. *American Journal of Public Health, 89*(11), 1728-1731.

Fiscella, K., & Kitzman, H. (2009). Disparities in academic achievement and health: The intersection of child education and health policy. *Pediatrics, 123*(3), 1073-1080.

Fraser, A. M., Brockert, J. E., & Ward, R. H. (1995). Association of young maternal age with adverse reproductive outcomes. *The New England Journal of Medicine, 332*(17), 1113-1118.

Fraser, E. D. G., Dougill, A. J., Mabee, W. E., Reed, M., & McAlpine, P. (2006). Bottom up and top down: Analysis of participatory processes for sustainability indicator identification as a pathway to community empowerment and sustainable environmental management. *Journal of Environmental Management, 78*(2), 114-127.

Freebairn, D. M., & King, C. A. (2003). Reflections on collectively working toward sustainability: indicators for indicators! *Animal Production Science*, *43*(3), 223-238.

Freudenberg, M. (2003). *Composite indicators of country performance: A critical assessment*. (2003/16). Paris, France: OECD Publishing.

Friedman, M. (2005). *Trying hard is not good enough*. Victoria, BC: Trafford Publishing.

Furr, R. M., & Bacharach, V. R. (2008). *Psychometrics: An introduction*. Thousand Oaks, CA: Sage.

Garnier, H. E., Stein, J. A., & Jacobs, J. K. (1997). The process of dropping out of high school: A 19-year perspective. *American Educational Research Journal, 34*(2), 395-419.

Gates, B. (2013). *Measuring progress*. Bill and Melinda Gates Foundation. Retrieved from http://www.gatesfoundation.org/who-we-are/resources-and-media/annual-letters-list/annual-letter-2013

Gaye, A. (2007). *Constructing composite indices*. Paper presented at the HDRO/RBA Regional Technical Workshop on Measuring Human Development, Nairobi. Retrived from http://www.powershow.com/view/134817-M2U1N/Constructing_composite_indices_Amie_Gaye_Policy_Specialist_UNDP_HDRO_powerpoint_ppt_presentation

Glaser, D. (2000). Child abuse and neglect and the brain - A review. *Journal of Child Psychology and Psychiatry, 41*(1), 97-116.

Goffreda, C. T., & DiPerna, J. C. (2010). An empirical review of psychometric evidence for the Dynamic Indicators of Basic Early Literacy Skills. *School Psychology Review, 39*(3), 463-483.

Good, R. H., & Kaminski, R. A. (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504-528.

Grace, J. B., & Bollen, K. A. (2008). Representing general theoretical concepts in structural equation models: The role of composite variables. *Environmental and Ecological Statistics, 15*(2), 191-213.

Gratz, Z. S. (2002). Review of the Basic Early Assessment of Reading. *Mental Measurements Yearbook, 16*.

Green, P. E. (2003). The undocumented: Educating the children of migrant workers in America. *Bilingual Research Journal, 27*(1), 51-71.

Greenland, S. (2001). Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. *International Journal of Epidemiology, 30*(6), 1343-1350.

Gutman, L. M., Sameroff, A. J., & Cole, R. (2003). Academic growth curve trajectories from 1st grade to 12th grade: Effects of multiple social risk factors and preschool child factors. *Developmental Psychology, 39*(4), 777-790.

Hair, E., Halle, T., Terry-Humen, E., Lavelle, B., & Calkins, J. (2006). Children's school readiness in the ECLS-K: Predictions to academic, health, and social outcomes in first grade. *Early Childhood Research Quarterly, 21*(4), 431-454.

Halle, T., Forry, N., Hair, E., Perper, K., Wander, L., & Vick, J. (2009). *Disparities in early learning and development: Lessons from the Early Childhood Longitudinal Study - Birth Cohort (ECLS-B)*. Washington, DC: Child Trends.

Hamilton, B. E., Martin, J. A., & Ventura, S. J. (2012) Births: Preliminary data for 2011. *National Vital Statistics Reports: Vol. 61*. Hyattsville, MD: National Center for Health Statistics.

Hammond, J. L. (1973). Two sources of error in ecological correlations. *American Sociological Review, 38*(6), 764-777.

Hardin, A., & Marcoulides, G. A. (2011). A commentary on the use of formative measurement. *Educational and Psychological Measurement, 71*(5), 753-764.

Heaviside, S., & Farris, S. (1993). *Public school kindergarten teachers' views on children's readiness for school*. Washington, DC: National Center for Education Statistics.

Helton, J. C., Johnson, J. D., Rollstin, J. A., Shiver, A. W., & Sprung, J. L. (1995). Uncertainty and sensitivity analysis of food pathway results with the MACCS reactor accident consequence model. *Reliability Engineering & System Safety, 49*(2), 109-144.

High, P. C. (2008). School readiness. *Pediatrics, 121*(4), e1008-e1015.

Hosp, J. L., & Reschly, D. L. (2004). Disproportionate representation of minority students in special education: Academic, demographic, and economic predictors. *Exceptional Children, 70*(2), 185-199.

Houweling, T., Kunst, A., & Mackenbach, J. (2003). Measuring health inequality among children in developing countries: does the choice of the indicator of economic status matter? *International Journal for Equity in Health, 2*(1), 1-12.

Hudrliková, L. (2013). Composite indicators as a useful tool for international comparison: The Europe 2020 example. *Prague Economic Papers, 2013*(4), 459-473.

Huston, A. C., & Bentley, A. C. (2010). Human development in societal context. *Annual review of psychology, 61*, 411-437.

Jacob, S. G., & Willits, F. K. (1994). Objective and subjective indicators of community evaluation: A Pennsylvania assessment. *Social Indicators Research, 32*(2), 161-177.

Jacobs, R., Smith, P., & Goddard, M. (2004). *Measuring performance: An examination of composite performance indicators*. (CHE Technical Paper Series 29). The University of York, Center for Health Economics. Retrieved from http://www.york.ac.uk/che/pdf/tp29.pdf

Johnson, K., Davidson, L. L., Theberge, S., & Knitzer, J. (2008) State indicators for early childhood. *Short Take No. 7*: National Center for Children in Poverty.

Jones, B., & Andrey, J. (2007). Vulnerability index construction: Methodological choices and their influence on identifying vulnerable neighborhoods. *International Journal of Emergency Management, 4*(2), 269-295.

Kaplan, D., & Walpole, S. (2005). A stage-sequential model of reading transitions: Evidence from the early childhood longitudinal study. *Journal of Educational Psychology, 97*(4), 551-563.

Kelley, S. J. (2003). Cumulative environmental risk in substance abusing women: early intervention, parenting stress, child abuse potential and child development. *Child Abuse & Neglect*, *27*(9), 993-995.

Kickham, K. (2000). *Work requirements and welfare participation: Analyzing the effects of the JOBS program* (Unpublished doctoral dissertation). University of Oklahoma, Norman, OK.

King, G. (1997). *A solution to the ecological inference problem: Reconstructing individual behavior from aggregate data*. Princeton, NJ: Princeton University Press.

Klerman, L. V. (1988). School absence -- a health perspective. *Pediatric Clinics of North America, 35*(6), 1253-1269.

Knitzer, J., & Lefkowitz, J. (2005) Resources to promote social and emotional health and school readiness in young children and families: A community guide. New York, NY: National Center for Children in Poverty.

Kramer, M. S. (2003). The epidemiology of adverse pregnancy outcomes: an overview. *The Journal of nutrition, 133*(5), 1592S-1596S.

Kramer, M. S., Séguin, L., Lydon, J., & Goulet, L. (2001). Socioeconomic disparities in pregnancy outcome: Why do the poor fare so poorly? *Paediatric and perinatal epidemiology, 14*(3), 194-210.

Lane, S. D., Webster, N. J., Levandowski, B. A., Rubinstein, R. A., Keefe, R. H., Wojtowycz, M. A., . . . Aubry, R. H. (2008). Environmental injustice: Childhood lead Poisoning, teen pregnancy, and tobacco. *Journal of Adolescent Health, 42*(1), 43-49.

Laosa, L. M., & Ainsworth, P. (2007) Is public pre-K preparing Hispanic children to succeed in school? *Preschool Policy Matters, Issue 13*. New Brunswick, NJ: NIEER.

Larson, H. A., El Ramahi, M. K., Conn, S. R., Estes, L. A., & Ghibellini, A. B. (2010). Reducing test anxiety among third grade students through the implementation of relaxation techniques. *Journal of School Counseling, 8*(19), n19.

Larson, K., Russ, S. A., Crall, J. J., & Halfon, N. (2008). Influence of multiple social risks on children's health. *Pediatrics, 121*(2), 337-344.

Lave, J. R., Keane, C. R., Lin, C. J., Ricci, E. M., Amersbach, G., & LaVallee, C. P. (1998). Impact of a children's health insurance program on newly enrolled children. *JAMA, 279*(22), 1820-1825.

Lazarte Alcalá, N., Salehezadeh, Z., & Schumacher, K. (2013). Oklahoma school readiness risk report 2013: Predictors in school readiness. *Practice and Policy Research Quarterly, 1*(2).

Lazarte Alcalá, N., & Schumacher, K. (2014). Oklahoma school readiness reach-by-risk report 2014. *Practice and Policy Research Quarterly, 2*(1).

Li, H., & Wu, J. (2006). Uncertainty analysis in ecological studies: an overview *Scaling and Uncertainty Analysis in Ecology* (pp. 45-66): Springer.

Literacy First framework. (2014). Retrieved May 27, 2014, from http://www.catapultlearning.com/professional-development-2/literacy-first/

Louisiana State University/Tulane Early Childhood Policy and Data Center. (2012). *Early childhood risk and reach in Louisiana*. Retrieved from http://www.gov.state.la.us/assets/2012_Risk_and_Reach_Report_online-correctedWEBSITE.pdf

Luman, E. T., McCauley, M. M., Shefer, A., & Chu, S. Y. (2003). Maternal characteristics associated With vaccination of young children. *Pediatrics, 111*(Supplement 1), 1215-1218.

Magnuson, K. (2003) The effect of increases in welfare mothers' education on their young children's academic and behavioral outcomes: Evidence from the national evaluation of Welfare-to-Work strategies child outcomes study. *Discussion Paper no. 1274-03*. Madison, WI: Institute for Research on Poverty.

Martin, J., Hamilton, B., Ventura, S., Osterman, M. J. K., Wilson, E. C., & Mathews, T. J. (2012). Births: Final data for 2010. *National Vital Statistics Reports, 61*(1).

Mashburn, A. J., & Pianta, R. C. (2006). Social Relationships and School Readiness. *Early Education and Development, 17*(1), 151-176.

Masten, A. S., & Coatsworth, J. D. (1998). The development of competence in favorable and unfavorable environments: Lessons from research on successful children. *American Psychologist, 53*(2), 205-220.

Masten, A. S., & Gewirtz, A. H. (2006). Vulnerabilty and resilience in early child development. In K. McCartney & D. Phillips (Eds.), *Blackwell handbook of eraly childhood development* (pp. 22-43).

Mather, M., & Dupuis, G. (2012). *The new KIDS COUNT index*. Baltimore, MD: Retrieved from http://www.aecf.org/m/resourcedoc/AECF-KIDSCOUNTIndex-2012.pdf

Mathur, S., & Parameswaran, G. (2011). Educating the migrant child: Professional development for teachers of young children of seasonal farm workers. *The Journal of Multiculturalism in Education, 7*.

Mayer, A. L. (2008). Strengths and weaknesses of common sustainability indices for multidimensional systems. *Environment International, 34*(2), 277-291.

McLafferty, S., & Grady, S. (2004). Prenatal care need and access: A GIS analysis. *Journal of Medical Systems, 28*(3), 321-333.

McLanahan, S. (2004). Diverging destinies: How children are faring under the second demographic transition. *Demography, 41*(4), 607-627.

McLanahan, S., & Sandefur, G. (1994). *Growing up with a single parent: What hurts, what helps*. Cambridge, MA: President and Fellows of Harvard College.

Meisels, S. J. (1999). Assessing readiness *The transition to kindergarten* (pp. 39-66). Baltimore: Brookes.

Messer, L., Laraia, B., Kaufman, J., Eyster, J., Holzman, C., Culhane, J., . . . O'Campo, P. (2006). The development of a standardized neighborhood deprivation index. *Journal of Urban Health, 83*(6), 1041-1062.

Moodie, S., & Rothenberg, L. (2011). *District of Columbia early childhood risk and reach assessment*. Washington, DC: Child Trends. Retrieved from http://osse.dc.gov/sites/default/files/dc/sites/osse/publication/attachments/ChildTrends_RiskReach_final%20(2).pdf

Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005) *Tools for composite indicators building*. European Commission Directorate-General Joint Research Centre.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2008). *Handbook on constructing composite indicators*. Paris: Organisation for Economic Co-operation and Development

National Center for Education Statistics. (2009). *Achievement gaps: How Black and White students in public schools perform in mathematics and reading on the National Assessment of Educational Progress*. (NCES 2009-455). Washington, DC.

National Center for Education Statistics. (2011). *Achievement gaps: How Hispanic and White students in public schools perform in mathematics and reading on the National Assessment of Educational Progress*. (NCES 2011-459). Washington, DC.

National Education Goals Panel. (1997). *Special early childhood report*. Washington, DC: US Government Printing Office Retrieved from http://govinfo.library.unt.edu/negp/reports/spcl.pdf

National Education Goals Panel.(1999) *The National Education Goals report: Building a nation of learners*. Retrieved from http://govinfo.library.unt.edu/negp/reports/99rpt.pdf

National Research Council and Institute of Medicine. (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academies Press.

Needell, B., Brookhart, M. A., & Lee, S. (2003). Black children and foster care placement in California. *Children and Youth Services Review, 25*(5–6), 393-408.

Nevárez-La Torre, A. A. (2012). Transiency in urban schools: Challenges and opportunities in educating ELLs with a migrant background. *Education and Urban Society, 44*(1), 3-34.

National Institute of Child Health and Human Development. (2002). Child-care structure → process → outcome: Direct and indirect effects of child-care quality on young children's development. *Psychological Science, 13*(3), 199-206.

Nunnaly, J. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.

Obama, B. H. (2013, February 12). *State of the Union Address*. Speech presented at Washington, D.C. Retrieved from http://www.whitehouse.gov/the-press-office/2013/02/13/fact-sheet-president-obama-s-plan-early-education-all-americans

Oklahoma Immunization Act, 70 Okla. Stat. Ann., Pub. L., §1210-191 Stat. (1970).

Parris, T. M., & Kates, R. W. (2003). Characterizing and measuring sustainable development. *Annual Review of Environment and Resources, 28*(1), 559-586.

Pears, K. C., Fisher, P. A., & Bronz, K. D. (2007). An intervention to promote social emotional school readiness in foster children: Preliminary outcomes from a pilot study. *School Psychology Review, 36*(4), 665-673.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction*. Belmont, CA: Wadsworth Publishing.

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Sharon Lynn, K., & Yazejian, N. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, *72*(5), 1534-1553.

Pelletier, J., & Brent, J. (2002). Parent participation in children' school readiness: The effects of parental self-efficacy, cultural diversity and teacher strategies. *International Journal of Early Childhood*, *34*(1), 45-60.

Pennsylvania Office of Child Development and Early Learning. (2012). *Pennsylvania program reach and risk assessment*. Retrieved from http://www.pakeys.org/uploadedContent/Docs/Reach_and_Risk_2010_2011_final_release_July_2012.pdf

Piotrkowski, C. S., Botsko, M., & Matthews, E. (2004). Parents' and teachers' beliefs about children's school readiness in a high-need community. *Early Childhood Research Quarterly, 15*(4), 537-558.

Rathvon, N. (2004). *Early reading assessment: A practioner's handbook*: The Guilford Press.

Reading Sufficiency Act - Programs of Reading Instruction, 70 Okla. Stat. §§1210.508C, Pub. L. (1997 & Supp. 2014).

Renger, R., & Titcomb, A. (2002). A three-step approach to teaching logic models. *American Journal of Evaluation*, *23*(4), 493-503.

Rhode Island KIDS COUNT. (2005). *Getting Ready*. Retrieved from http://gettingready.org/matriarch/OnePiecePage.asp_Q_PageID_E_93_A_PageName_E_whatwedoindic

Rimm-Kaufman, S. E., & Pianta, R. C. (2000). An Ecological Perspective on the Transition to Kindergarten: A Theoretical Framework to Guide Empirical Research. *Journal of Applied Developmental Psychology, 21*(5), 491-511.

Robinson, C. A., Sepe, S. J., & Lin, K. F. Y. (1993). The president's child immunization initiative - A summary of the problem and the response. *Public Health Reports, 108*(4), 419-425.

Robinson, W. S. (2009). Ecological correlations and the behavior of individuals. *International Journal of Epidemiology, 38*(2), 337-341.

Rodriguez, E. T., & Tamis-LeMonda, C. S. (2011). Trajectories of the home learning environment across the first 5 years: Associations with children's vocabulary and literacy skills at prekindergarten. *Child Development, 82*(4), 1058-1075.

Rossi, R. J., & Gilmartin, K. J. (1980). *The handbook of social indicators: sources, characteristics, and analysis*: Garland STPM Press New York.

Rouse, H. L., & Fantuzzo, J. W. (2009). Multiple risks and educational well being: A population-based investigation of threats to early school success. *Early Childhood Research Quarterly, 24*(1), 1-14.

Roy, R., Chan, N. W., & Rainis, R. (2014). Rice farming sustainability assessment in Bangladesh. *Sustainability science, 9*(1), 31-44.

Roy, S., Tarafdar, M., Ragu-Nathan, T., & Marsillac, E. (2012). The effect of misspecification of reflective and formative constructs in operations and manufacturing management research. *The Electronic Journal of Business Research Methods, 10*(1), 34-52.

Saisana, M. (2008). *2007 Composite Learning Index: Robustness issues and critical assessment*. (EUR 23274 EN 2008). Luxembourg: European Commission. Retrieved from http://www.ccl-cca.ca/pdfs/CLI/2007/JRC_CLI2007ReportonValidation_Finalpdf.pdf

Saisana, M. (2013, Nov. 12). *Michaela Saisana on EC guidelines to measure sustainable development* [Video file]. FEEM International Conference on Methodologies and Indicators for Green Growth Measurement. Retrieved from https://www.youtube.com/watch?v=NXbQqMXgPMY

Saisana, M., & Munda, G. (2008). *Knowledge economy: Measures and drivers*. (EUR 23486 EN). Luxembourg: JRC Scientific and Technical Reports.

Saisana, M., & Saltelli, A. (2008). Expert Panel Opinion and Global Sensitivity Analysis for Composite Indicators. In F. Graziani (Ed.), *Computational Methods in Transport: Verification and Validation* (Vol. 62, pp. 251-275): Springer Berlin Heidelberg.

Saisana, M., Tarantola, S., & Saltelli, A. (2005). Uncertainty and sensitivity techniques as tools for the analysis and validation of composite indicators. *Journal of the Royal Statistical Society A, 168*(2), 1-17.

Saltelli, A., Munda, G., & Nardo, M. (2006). From complexity to multidimensionality: The role of composite indicators for advocacy of EU reform. *Tijdschrift voor Economie en Management, LI*(3), 221-235.

Saltelli, A., Nardo, M., Saisana, M., & Tarantola, S. (2005). Composite indicators - The controversy and the way forward. In E. Giovannini (Ed.), *Statistics, knowedge* *and policy: Key indicators to inform decision making* (pp. 359-371). Paris: Organisation for Economic Co-operation and Development (OECD).

Saluja, G., Early, D. M., & Clifford, R. M. (2002). Demographic characteristics of early childhood teachers and structural elements of early care and education in the United States. *Early Childhood Research and Practice, 4*(1).

Salzinger, S., Feldman, R. S., Stockhammer, T., & Hood, J. (2002). An ecological framework for understanding risk for exposure to community violence and the effects of exposure on children and adolescents. *Aggression and Violent Behavior, 7*(5), 423-451.

Sameroff, A. J., & Fiese, B. H. (2000). Transactional regulation: The developmental ecology of early intervention. In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (pp. 135-159). New York: Cambridge University Press.

Sapolsky, R. M. (1996). Why stress is bad for your brain. *Science*, *273*(5276), 749-750.

Sargent, J. D., Brown, M. J., Freeman, J. L., Bailey, A., Goodman, D., & Freeman, D. H. (1995). Childhood lead poisoning in Massachusetts communities: Its association with sociodemographic and housing characteristics. *American Journal of Public Health*, *85*(4), 528-534.

Scaglion, R., & Condon, R. G. (1980). Determinants of attitudes toward city police. *Criminology*, *17*(4), 485-494.

Schmidtlein, M. C., Deutsch, R. C., Piegorsch, W. W., & Cutter, S. L. (2008). A sensitivity analysis of the Social Vulnerability Index. *Risk Analysis*, *28*(4), 1099-1114.

Schulman, K., & Barnett, W. S. (2005). *The benefits of prekindergarten or middle-income children*. New Brunswick, NJ: NIEER.

Sektnan, M., McClelland, M. M., Acock, A., & Morrison, F. J. (2010). Relations between early family risk, children's behavioral regulation, and academic achievement. *Early Childhood Research Quarterly, 25*(4), 464-479.

Shanahan, T. (2003). Review of the DIBELS: Dynamic Indicators of Early Literacy Skills, sixth edition. *Mental Measurements Yearbook, 16*.

Sharpe, A., & Andrews, B. (2012). *An assessment of weighting methodologies for composite indicators: The case of the index of economic well-being*. (CSLS Research Report No. 2012-10). Ottawa, Ontario: Centre for the Study of Living Standards. Retrieved from http://www.csls.ca/reports/csls2012-10.pdf

Sheldon, E. B., & Parke, R. (1975). Social indicators. *Science, 188*(4189), 693-698.

Shonkoff, J. P., & Garner, A. S. (2012). The lifelong effects of early childhood adversity and toxic stress. *Pediatrics, 129*(1), e232-e246.

Simpson, D. M. (2006). Indicator issues and proposed framework for a disaster preparedness index (DPi). University of Louisville: Center for Hazards Research and Policy Development for the Fritz Institute.

SmithBattle, L. (2007). “I wanna have a good future”: Teen mothers' rise in educational aspirations, competing demands, and limited school support. *Youth & Society, 38*(3), 348-371.

Snow, K. L. (2006). Measuring school readiness: Conceptual and practical considerations. *Early Education and Development, 17*(1), 7-41.

Stanton-Chapman, T. L., Chapman, D. A., Kaiser, A. P., & Hancock, T. B. (2004). Cumulative risk and low-income children's language development. *Topics in Early Childhood Special Education, 24*(4), 227-237.

Stovall-McClough, K. C., & Dozier, M. (2004). Forming attachments in foster care: Infant attachment behaviors during the first 2 months of placement. *Development and Psychopathology, 16*(2), 253-271.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Upper Saddle River, NJ: Pearson Allyn & Bacon.

Tallis, K. (2005). Social indicator systems - Foundations for policy design and evaluation? In E. Giovannini (Ed.), *Statistics, knowledge adn policy: Key indicators to inform decision making* (pp. 198-216). Paris: OECD.

Tate, E. (2012). Social vulnerability indices: a comparative assessment using uncertainty and sensitivity analysis. *Natural Hazards, 63*(2), 325-347.

Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. London: Falmer Press.

Thivierge, M.-N., Parent, D., Bélanger, V., Angers, D. A., Allard, G., Pellerin, D., & Vanasse, A. (2014). Environmental sustainability indicators for cash-crop farms in Quebec, Canada: A participatory approach. *Ecological Indicators, 45*, 677-686.

Thomas, D. V., Seshadri, R., Cesarone, B., Fowler, S., Goerge, R., & Danner, N. (2012). Assessing risk: Descriptive data related to risk factors experienced by young children and their families in Illinois. *Illinois Early Childhood Asset Map and Chapin Hall*.

Thompson, B. L., Levitt, P., & Stanwood, G. D. (2009). Prenatal exposure to drugs: Effects on brain development and implications for policy and education. *Nature Reviews Neuroscience, 10*, 303-312.

Tong, V. T., Dietz, P. M., Morrow, B., D'Angelo, D. V., Farr, S. L., Rockhill, K. M., & England, L. J. (2013). *Trends in smoking before, during, and after pregnancy - Pregnancy Risk Assessment Monitoring System, United States, 40 sites, 2000-2010*. (62(SS06); 1-19). Centers for Disease Control and Prevention. Retrieved from http://www.cdc.gov/mmwr/preview/mmwrhtml/ss6206a1.htm

Tucker, C. M., Zayco, R. A., Herman, K. C., & Reinke, W. M. (2002). Teacher and child variables as predictors of academic engagement among low-income African-American children *Psychology in the Schools, 39*(4), 477-488.

United Nations Development Programme. (1990). *Concept and measurement of human development*. Retrieved from http://hdr.undp.org/sites/default/files/reports/219/hdr_1990_en_complete_nostats.pdf

United Nations Development Programme. (2013). *The rise of the south: Human progress in a diverse world*. Retrieved from http://hdr.undp.org/sites/default/files/reports/14/hdr2013_en_complete.pdf

U.S. Census Bureau, U.S. Department of Commerce, Economics and Statistics Administration. (2014). *The undercount of young children*. Retrieved from http://www.census.gov/library/working-papers/2014/demo/2014-undercount-children.html

Vernon-Feagans, L., & Blair, C. (2006). Measurement of school readiness. *Early Education and Development*, *17*(1), 1-5.

Vernon-Feagans, L., Odom, E., Pancsofar, N., & Kainz, K. (2008). Farkas and Hibel, and a transactional/ecological model of readiness and inequality. In A. Booth & A. C. Crouter (Eds.), *Disparities in school readiness: How families contribute to transitions into school* (pp. 61-78). New York: Lawrence Erlbaum Associates.

West, J., Denton, K., & Germino-Hausken, E. (2000). America's kindergartners: Findings from the early childhood longitudinal study, kindergarten class of 1998-99: Fall 1998. *Education Statistics Quarterly, 2*(1), 7-13.

West, J., Germino-Hausken, E., & Collins, M. (1995). *Readiness for kindergarten: Parent and teacher beliefs*. (Statistics in brief). Washington, DC: National Center for Education Statistics.

Widom, C. S. (2000). Childhood victimization: Early adversity, later psychopathology. *National Institute of Justice Journal, 242*, 3-9.

Wong, C. (2006). *Indicators for urban and regional planning: The interplay of policy and methods*. New York, NY: Routledge.

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (Eds.). (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage.

Zill, N., & West, J. (2001) Entering kindergarten: A portrait of American children when they begin school. *Findings from the Condition of Education, 2000*. Washington, DC: National Center for Education Statistics.

Oklahoma State University Institutional Review Board

Request for Determination of Non-Research or Non-Human Subject

Federal regulations and OSU policy require IRB review of all research involving human subjects. Some categories of research are difficult to discern as to whether they qualify as human subject research. Therefore, the IRB has established policies and procedures to assist in this determination.

1. Principal Investigator Information

First Name: Krista		Middle Initial: S	Last Name: Schumacher
Department/Division: SES/REMS		College: Education	
Campus Address:		Zip+4:	
Campus Phone:	Fax:	Email: krista.schumacher@okstate.edu	
Complete if PI does not have campus address:			
Address: 33721 W. McCrackin Rd.		City: Mannford	
State: OK	Zip: 74044	Phone: 918-284-7276	

2. Faculty Advisor **(complete if PI is a student, resident, or fellow)** ☐ **NA**

Faculty Advisor's name: Jam Khojasteh		Title: Assistant Professor
Department/Division: SES/REMS		College: Education
Campus Address: OSU-Tulsa, MH2444A		Zip+4:
Campus Phone: 918-594-8226	Fax:	Email: jam.khojasteh@okstate.edu

3. Study Information:

A. **Title**
Informing Early Childhood Policy: An Analysis of the Sensitivity of a School Readiness Risk Index to Changes in Indicator Selection

B. **Give a brief summary of the project. (See instructions for guidance)**
The overarching research question to be answered is how changes to a set of county-level indicators alter index results and their relationship with a proxy measure of school readiness. Examples of indicators include rates of child poverty, teen pregnancies, single parents, and prenatal care. A total of 16 such indicators will be used, in addition to county-level percent of kindergarteners scoring below benchmark on legislatively mandated literacy assessments administered in Oklahoma schools. Data on 10 indicators have been published in the Oklahoma Department of Human Services (DHS) publication, *Oklahoma School Readiness Reach-by-Risk Report 2014*. Data for remaining indicators are available through OK2SHARE, an online database of the Oklahoma Department of Health, and through open records requests for data not available online. All data used in this study is available to the general public.

C. **Describe the subject population/type of data/specimens to be studied. (See instructions for guidance)**
The unit of analysis is counties, and all data will be at the aggregate level. No individual-level data is being collected; hence, no identifiers are used. It is impossible for any individual to be linked to the data. The type of data is administrative and census data ordinarily collected by state and federal agencies.

Oklahoma State University Institutional Review Board

Request for Determination of Non-Research or Non-Human Subject

4. **Determination of "Research".**
   One of the following must be "no" to qualify as "non-research":

   A. Will the data/specimen(s) be obtained in a systematic manner?
   ☒ No ☐ Yes

   B. Will the intent of the data/specimen collection be for the purpose of contributing to generalizable knowledge (the results (or conclusions) of the activity are intended to be extended beyond a single individual or an internal program, i.e. widely or universally applicable)?
   ☒ No ☐ Yes

5. **Determination of "Human Subject".**

   A. Does the research involve obtaining information about living individuals?
   ☐ No ☒ Yes
   If no, then research does not involve human subjects, <u>no other information is required</u>.
   If yes, proceed to the following questions.

   All of the following must be "no" to qualify as "non-human subject":

   B. Does the study involve intervention or interaction with a "human subject"?
   ☒ No ☐ Yes

   C. Does the study involve access to identifiable private information?
   ☒ No ☐ Yes

   D. Are data/specimens <u>received</u> by the Investigator with identifiable private information?
   ☒ No ☐ Yes

   E. Are the data/specimen(s) coded such that a link exists that could allow the data/specimen(s) to be re-identified?
   ☒ No ☐ Yes
   If "Yes," is there a written agreement that prohibits the PI and his/her staff access to the link?
   ☐ No ☐ Yes

6. **Signatures**

   Signature of PI Krista Schumacher **Date** 10-15-14

   Signature of Faculty Advisor ____________ **Date** 10-15-14
   (If PI is a student)

☒ Based on the information provided, the OSU-Stillwater IRB has determined that this project **does not** qualify as human subject research as defined in 45 CFR 46.102(d) and (f) and **is not subject to oversight by the OSU IRB.**

☐ Based on the information provided, the OSU-Stillwater IRB has determined that this research **does** qualify as human subject research and **submission of an application for review by the IRB is required.**

______________________ 1-7-15
Dr. Hugh Crethar, IRB Chair Date

VITA

Krista Sue Schumacher

Candidate for the Degree of

Doctor of Philosophy

Thesis: INFORMING EARLY CHILDHOOD POLICY: AN ANALYSIS OF THE SENSITIVITY OF A SCHOOL READINESS RISK INDEX TO CHANGES IN INDICATOR SELECTION

Major Field: Educational Psychology: Research and Evaluation

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Educational Psychology: Research and Evaluation at Oklahoma State University, Tulsa, Oklahoma in May, 2015.

Completed the requirements for the Master of Arts in Liberal Studies at University of Oklahoma, Norman, Oklahoma in May 2007.

Completed the requirements for the Bachelor of Science in Journalism at Oklahoma State University, Stillwater, Oklahoma in 1994.

Experience:

Senior Researcher, Office of Planning, Research and Statistics, Oklahoma Department of Human Services, 2013 to present

Associate, JCCI Resource Development Services, 2007 to present

Research and Teaching Assistant, Research, Evaluation, Measurement, and Statistics, Oklahoma State University, 2012 to 2014

Director, Office of Grant Development, Tulsa Community College, 2004 to 2006

Professional Memberships:

American Evaluation Association

Scholars Strategy Network