


March 2020

Predictive Validity of Standards-based and Curriculum-embedded Assessments for Predicting Readiness at Kindergarten Entry

Elizabeth Ashton DeCamilla
University of South Florida

Follow this and additional works at: <https://scholarcommons.usf.edu/etd>

 Part of the [Educational Assessment, Evaluation, and Research Commons](#), [Pre-Elementary, Early Childhood, Kindergarten Teacher Education Commons](#), and the [Statistics and Probability Commons](#)

Scholar Commons Citation

DeCamilla, Elizabeth Ashton, "Predictive Validity of Standards-based and Curriculum-embedded Assessments for Predicting Readiness at Kindergarten Entry" (2020). *Graduate Theses and Dissertations*. <https://scholarcommons.usf.edu/etd/8179>

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

Predictive Validity of Standards-based and Curriculum-embedded Assessments for Predicting
Readiness at Kindergarten Entry

by

Elizabeth Ashton DeCamilla

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Curriculum and
Instruction with a concentration in
Measurement and Evaluation
Department of Educational and Psychological
Studies
College of Education
University of South Florida

Major Professor: Robert F. Dedrick, Ph.D.
John Ferron, Ph.D.
Yi-Hsin Chen, Ph.D.
Danielle V. Dennis, Ph.D.

Date of Approval:
February 24, 2020

Keywords: Early Childhood Education, Kindergarten Readiness, Kindergarten Entry
Assessment, Predictive Validity, School Readiness

Copyright © 2020, Elizabeth Ashton DeCamilla

Dedication

I would like to dedicate this dissertation to my family and friends that have joined me along my entire educational journey. My parents and grandparents continually reminded me that education was a gateway not only to knowledge, but experiences that would not only benefit myself, but enhance the community and future generations. Secondly, to my husband David, who was a steadfast support through this entire process, often being the gentle guide when it was difficult to see the light at the end of the tunnel. My friends, both near and far, deserve special thanks for being my cheerleaders, proof-readers, and therapists. Lastly and most importantly, I dedicate this dissertation to my son, Drew Ashton DeCamilla. I hope I showed Drew that as long as an Ashton DeCamilla puts their mind to any goal, no matter how lofty, the goal WILL be achieved. I also hope I've shown Drew through this process that it is not just the goal that is important, but achieving the goal with strength, dignity, ethics and pride. No matter how long and challenging the journey that he or I decide to embark upon, an Ashton DeCamilla will not quit, but will adapt and complete with our heads held high. Lastly, an enormous level of pride and support to early childhood educators, especially Head Start teachers that continue to proud foundational school readiness skills for our youngest learners.

Acknowledgements

I would like to acknowledge Dr. Lori R. Wiggins for being my writing partner. Although we were in different doctoral programs, we spent many Saturday and Sundays together working on our dissertation proposals. We shed many tears, snacks and celebrations through this entire process. Her feedback and support have been unwavering.

Most importantly, enormous appreciation and gratitude is shared with my remarkable committee. Throughout the years, I have been lucky enough to have learned immensely from them within their feedback through each stage of the proposal and dissertation. A special thank you to Dr. Elizabeth Shaunessey-Dedrick for stepping in as my outside chair. Dr. Yi-Hsin Chen, thank you for the supports of looking at multi-level data through a new lens. To Dr. Danielle Dennis, thank you for opening my eyes slightly wider to the importance of continual research related to the policies that impact children during their educational experiences, especially unintended consequences impacting underrepresented populations. An immense and heart-felt thank you to Dr. John Ferron. Without learning from Dr. Ferron's vast statistical knowledge in the master's and doctoral program, I would not be currently working in the field nor have been able to complete this process. Most importantly, words can not express the level of gratitude and admiration I have for my major professor, Dr. Robert Dedrick. As a professor, his measurement and multi-level modeling courses were engaging, challenging and provided me with the foundations needed to be successful within the work force. As a major professor, he was always a strong instructional, moral and emotional support throughout the dissertation process.

TABLE OF CONTENTS

List of Tables	iii
Abstract	vi
CHAPTER ONE: INTRODUCTION.....	1
National Context	3
State Context.....	6
Statement of the Problem.....	8
Validity of Kindergarten Entry Assessments.....	9
Purpose of the Study	11
Research Questions.....	13
Limitations of the Study.....	14
Delimitations of the Study	14
Definition of Terms.....	15
Importance of the Study.....	16
Organization of the Study	17
CHAPTER TWO: REVIEW OF LITERATURE.....	18
Background Information of Early Childhood Education.....	19
Early History	21
Emergence of Universal Prekindergarten	24
Universal Prekindergarten in Florida.....	25
Early Childhood Learning Standards and Assessments.....	26
Overview.....	26
Good Start, Grow Smart Initiative	28
Head Start.....	31
Voluntary Prekindergarten in Florida	35
School Readiness at Kindergarten Entry	38
Validity	41
Overview.....	41
Predictive Validity	42
Present Study	43
CHAPTER THREE: METHOD	46
Purpose of the Study	46
Research Questions.....	46
Research Design.....	47
Data Sources	48
Florida VPK Assessment Measures.....	48
Teaching Strategies Gold Assessment	52
Work Sampling System P4, 5 th Edition	56

Measure Procedures	61
Participant Selection	62
Data Analysis	66
Variables	66
Preliminary Analysis.....	67
Statistical Analysis.....	67
Chapter Summary	70
CHAPTER FOUR: RESULTS	71
Data Collection	71
Descriptive Statistics.....	72
Group Statistics.....	81
Model Results	85
Models Examining Scales Independently	89
Models with All Subscales Included.....	91
Chapter Summary	94
CHAPTER FIVE: DISCUSSION.....	95
Review of Sample and Methods	96
Findings.....	98
Models Examining Scales Independently	99
Florida VPK Assessment Measures.....	99
Teaching Strategies Gold Assessment	100
Models with All Subscales Included.....	102
Limitations of the Study.....	103
Implications for Future Research and Policy	109
Conclusion	112
References	115
Appendix	122
Appendix A: History of Head Start Funding and Implementation	123
Appendix B: Work Sampling System P4, 5 th Edition Scoring Tables.....	137
Appendix C: Sample of Mplus Code of Huber-White Sandwich Estimator	138
Appendix D: Ordinal Logistic Regression Model Summary Tables	139
Appendix E: Approval from School District Review Board	155
Appendix F: Letter of Approval from University Institutional Review Board	156

List of Tables

Table 1:	Internal Consistency Reliability Ranges for VPK Assessment Measures (AP3) for Samples Used in Different Phases of VPK Measure Development	50
Table 2:	Florida VPK Assessment General Outcomes Subscale Variables for Kindergarten Readiness	53
Table 3:	Pearson Correlation Ranges between External Measures and Teaching Strategies GOLD® scale scores	55
Table 4:	Teaching Strategies GOLD® Curriculum-embedded Subscale Variables for Kindergarten Readiness	57
Table 5a:	Distribution of ECE Children Nested Within Classrooms Within Schools with Completed VPK Assessments.....	64
Table 5b:	Distribution of ECE Children Nested Within Classrooms Within Schools with Completed Teaching Strategies GOLD Assessments.....	65
Table 6:	Description of Independent Variables in the Multiple Logistic Regression Models.....	68
Table 7:	Preliminary Demographic Information of the Sample Cohorts.....	73
Table 8:	Descriptive Statistics: Full Sample for Cohort One	74
Table 9:	Descriptive Statistics: Gender Sample for Cohort One	75
Table 10:	Descriptive Statistics: Ethnicity Sample for Cohort One	76
Table 11:	Sample Sizes for Work Sampling System (WSS) Categorical Variables by Main Effects for Cohort One.....	77
Table 12:	Descriptive Statistics: Full Sample for Cohort Two.....	78
Table 13:	Descriptive Statistics: Gender Sample for Cohort Two.....	79

Table 14:	Descriptive Statistics: Ethnicity Sample for Cohort Two.....	80
Table 15:	Sample Sizes for Work Sampling System (WSS) Categorical Variables by Main Effects for Cohort Two	81
Table 16:	Descriptive Statistics and Effect Sizes by Gender Based on Main Effects – VPK Assessment Subscales – Cohort One	82
Table 17:	Descriptive Statistics and Effect Sizes by Ethnicity Based on Main Effects – VPK Assessment Subscales – Cohort One	82
Table 18:	Descriptive Statistics and Effect Sizes by Gender Based on Main Effects – Teaching Strategies GOLD Subscales – Cohort One	83
Table 19:	Descriptive Statistics and Effect Sizes by Ethnicity Based on Main Effects – Teaching Strategies GOLD Subscales – Cohort One	83
Table 20:	Descriptive Statistics and Effect Sizes by Gender Based on Main Effects – VPK Assessment Subscales – Cohort Two	84
Table 21:	Descriptive Statistics and Effect Sizes by Ethnicity Based on Main Effects – VPK Assessment Subscales – Cohort Two	84
Table 22:	Descriptive Statistics and Effect Sizes by Gender Based on Main Effects – Teaching Strategies GOLD Subscales – Cohort Two.....	85
Table 23:	Descriptive Statistics and Effect Sizes by Ethnicity Based on Main Effects – Teaching Strategies GOLD Subscales – Cohort Two.....	85
Table 24:	Summary of Predictor and Interaction Relationship by Each Assessment Measure Subscale	90
Table 25:	Summary of Predictor and Interaction Relationship by Assessment Measures with All Subscales Included.....	93
Table A1:	Logistic Models Predicting Kindergarten Readiness with VPK Assessments– Cohort 1	139

Table A2:	Logistic Models Predicting Kindergarten Readiness with VPK Assessment- Cohort 1	141
Table A3:	Logistic Models Predicting Kindergarten Readiness with VPK Assessments- Cohort 2	143
Table A4:	Logistic Models Predicting Kindergarten Readiness with VPK Assessments- Cohort 2	145
Table A5:	Logistic Models Predicting Kindergarten Readiness with Teaching Strategies GOLD® - Cohort 1	147
Table A6:	Logistic Models Predicting Kindergarten Readiness with Teaching Strategies GOLD® - Cohort 1	149
Table A7:	Logistic Models Predicting Kindergarten Readiness with Teaching Strategies GOLD® - Cohort 2	151
Table A8:	Logistic Models Predicting Kindergarten Readiness with Teaching Strategies GOLD® - Cohort 2	153

Abstract

As with traditional K-12 educational settings, early childhood assessments have been a primary source of information determining whether early educational experiences have promoted children's readiness to start school in kindergarten. The level of use of Kindergarten Entry Assessments (KEAs) has become more wide-spread to establish levels of school readiness at kindergarten entry.

This quantitative, correlational study of children in schools that have blended Head Start/Voluntary Prekindergarten funded programs examined the predictive relationships between the independent variables (i.e., VPK Assessments and Teaching Strategies GOLD®) and the dependent variable of kindergarten readiness, as measured by the Work Sampling System™ (WSS). Additionally, the study examined whether gender and ethnicity moderated the predictive relationships between the independent variables and kindergarten readiness.

Data from two cohorts of children enrolled in a blended Head Start/VPK funded program in 2014-2015 ($N = 604$) and 2015-2016 ($N = 565$) nested within 39 classrooms nested within 22 schools were analyzed using multiple ordinal logistic regressions to determine the predictive relationships between the four VPK Assessment subscales and the four Teaching Strategies GOLD® subscales. Analyses began by looking at the predictive relationships of the VPK Assessment and Teaching Strategies GOLD® subscales one subscale at a time. Next, a combined model of the four subscales of the VPK Assessment was examined followed by a combined model of the four subscales of the Teaching Strategies GOLD®. When examining each subscale predictor individually, a robust predictive relationship (i.e., a significant

relationship at $p < .01$ was observed in both cohorts) was shown for all subscales of the VPK Assessment (Print Knowledge, Phonological Awareness, Mathematics, and Oral Language/Vocabulary). For Teaching Strategies GOLD® all subscales except for Cognitive indicated a robust predictive relationship (Social-Emotional, Literacy, and Mathematics). However, when looking at the predictive relationships with all subscales in the model for the VPK Assessment, none of the four subscales had a robust predictive relationship. Similar results were found for the combined model with the four subscales of the Teaching Strategies GOLD®. Within all models, gender and ethnicity did not have robust moderating effects on the predictive relationships of VPK Assessments and Teaching Strategies GOLD®. These results indicated no evidence of subgroup differences with each scale of the VPK Assessment and Teaching Strategies GOLD®, providing one source of evidence of fairness of both measures. Implications related to these findings related to the predictive validity of these early childhood assessments on levels of kindergarten readiness are discussed.

CHAPTER ONE

INTRODUCTION

Early learning and development opportunities in the United States (US) have the potential to produce positive and potentially lasting impact on children. In order to meet federal grant competitions within the new millennium, US early childhood education has been experiencing change as many states have begun to revamp their educational systems, especially for children birth through kindergarten entry age. An increased emphasis on school readiness at kindergarten entry has placed a direct spotlight on assessment and accountability for early education programs. This increased focus has highlighted a need for psychometricians and educational researchers to prioritize partnerships with the early childhood community to ensure quality kindergarten entry assessment instruments are aligned with early learning standards and validated to indicate levels of school readiness (Goldstein & Flake, 2016). Early childhood experts encourage assessments of the whole child (National Research Council, 2008). However, valid and reliable early childhood and kindergarten entry assessment instruments are not readily available for all domains related to the whole child, in particular social and emotional development. This situation has provided the impetus for psychometricians, early education researchers, and state education agencies to develop valid early childhood assessments that have the potential to predict school readiness at kindergarten entry.

Within the last decade, the federal government has provided competitive funding opportunities focused on assessment practices focused on young children in early educational programs. State education agencies had the ability to compete for the Enhanced Assessment Grant and Race to the Top Early Learning Challenge Grant. The purpose of the Enhanced

Assessment Grant (2013) was to increase the quality of assessment instruments in elementary and secondary schools. In particular, the focus of the Enhanced Assessment Grant was for state agencies to develop or enhance a kindergarten entry assessment (KEA) aligned with a set of early learning standards and development. In September of 2013, the U.S. Department of Education awarded more than \$15 million in Enhanced Assessment Grants to three state education agencies, North Carolina, Maryland, and Texas.

As of January 2014, over 1 billion dollars in federal Race to the Top Early Learning Challenge grants were awarded to 20 states. A primary focus of the grants was for states to afford more opportunities for low-income and disadvantaged children to have access to high-quality early childhood experiences, and ensure that any use of assessment adheres to the recommendations of the Nation Research Council's report on early childhood. Within this 2008 report, the National Research Council defined a comprehensive early childhood system of developmental screening measures, formative assessments, measures of environmental quality, measures of the quality of adult-child interactions, and concluding with kindergarten entry assessment to determine school readiness. Goldstein and Flake (2016) stated that there is a national need for developmentally appropriate, psychometrically sound instruments to monitor young children's learning and development that can be used to evaluate the effectiveness of their early childhood educational programs. This fueled the need for federal investments. Although the state of Florida was not awarded either an Enhanced Assessment grant or Race to the Top Early Learning Grant, a kindergarten entry assessment has been in place for all incoming kindergarten students in public and charter schools since 2001.

National Context.

As with traditional K-12 educational settings, early childhood assessments have been a primary source of information determining whether early educational experiences have promoted children's readiness to start school in kindergarten. As the continuum of standards within the early childhood community has been revised to include a cognitive and academic focus at the turn of the 21st century, the level of use of standards-based assessment practices has become more wide-spread not only as an assessment practice, but as a method to predict continued academic school readiness at kindergarten entry. This assessment process for predicting kindergarten readiness made sense to practitioners and policy makers as the assessment process mirrored the academically-focused early childhood standards, as well as traditional, norm-referenced assessment practices within the kindergarten through twelfth grade (K-12) educational system.

However, over the past decade, as early childhood standards began to be revised to contain all areas of young children's development, so did the daily assessment practices within early childhood settings. A literature review of early childhood assessment practices conducted in 2014 by the Administration of Children and Families Office of Planning, Research and Evaluation (OPRE) found that although early childhood teachers continue to utilize standards-based general outcome measures per federal or state requirements to predict school readiness at kindergarten entry, the use of ongoing curriculum-embedded assessment measures practices to predict future success and inform direct, individualized instruction are more prevalent. The authors found, "ongoing assessment practices are used by early childhood teachers to adjust instructional or developmental practices and content to better meet the individual strengths, needs, and interests of young children" (Akers et al., 2014, p. 1). The National Center of

Quality Teaching and Learning (2012, p. 2) operationally defined ongoing assessment practices as the process of “continuing observation and documentation teachers complete to determine whether teaching practices need to be adapted to better meet children’s needs.”

Although early childhood standards and assessment practices have evolved toward more curriculum-aligned performance assessments at the district and school level over the past decade, metrics to determine school readiness at kindergarten entry have not been quick to follow (Harvey & Ohle, 2018). Prompted by legislation and funding opportunities (e.g., Race for the Top-Early Learning Challenge and Enhanced Assessment Grants), the number of states developing, piloting or implementing kindergarten entry assessments (KEA) have been on the rise. The purpose of a KEA is to document a child’s developmental skills and behaviors that have been shown to predict long term student success (U.S. Department of Education, 2001; 2014; Harvey & Ohle, 2018). In 2010, the Center on Enhancing Early Learning Outcomes (CEELO) noted that only seven states utilized KEA for aggregating the number and percentage of children presenting school readiness at the district and state data levels (Alaska, Connecticut, Florida, Hawaii, Maryland, Minnesota, and Vermont). By 2012, the number of states requiring assessments rose to 25 with 12 states assessing children at kindergarten entry, 10 during the school year and three at both entry and during the year (Connors-Tadros, 2014). In 2013, as part of the Enhanced Assessment Grant (EAG) program, the competition made awards that supported the development or enhancement of a kindergarten entry assessment aligned to a set of early learning and development standards. Competitive preference priority 1 also focused on early learning collaboration efforts among states (U.S. Department of Education, 2013). A state educational agency (SEA) or a consortium of SEAs, as defined in section 9101 (41) of the Elementary and Secondary Education Act of 1965, as amended (ESEA) had the ability to apply

for an EAG with an estimated award size of \$4,600,000. Thirty-four states described plans for a KEA in their Race to the Top – Early Learning Challenge applications with only three states receiving the Enhanced Assessment Grant award: Maryland, North Carolina, and Texas.

Although an exponential growth of KEAs within the past ten years, researchers have found many complexities regarding the uniformity, utility, and validity of these kindergarten entry assessments (Goldstein & Flake, 2016; Pierson, 2018; Harvey & Ohle, 2018). Harvey and Ohle stated that although state KEAs vary in terms of the instruments used and the types of learning assessed, the movement toward aligning early childhood systems with K-12 education with the KEAs being a vital component to the systems' success. The purpose of the Enhanced Assessment Grant was to support the development or enhancement of a kindergarten entry assessments to provide state agencies, teachers and parents valid and reliable information on each child's development across defined essential domains of school readiness. These essential school readiness domains are: (1) language and literacy development; (2) cognition and general knowledge (including early mathematics and early scientific development); (3) approaches toward learning; (4) physical well-being and motor development (including adaptive skills); and (5) social and emotional development. These domains must also be aligned with the applying states' K-3 academic content standards in, at a minimum, early literacy and mathematics (U.S. Department of Education, 2013). The KEA was to be developed to include all students, including English learners and students with disabilities and data to support educators in providing effective learning opportunities to each child to in turn, help close the achievement gaps. The KEA must not be used to prevent children's entry into kindergarten.

In 2017, the Center on Enhancing Early Learning Outcomes revisited the implementation of states' efforts in implementing KEAs. States continue to be at differing stages of KEA

implementation that range from statewide implementation of a standardized assessment measure to piloting a tool or just beginning the conversation (Weisenfeld, 2017). In the 2011-2012 NIEER Yearbook, 19 states had an established kindergarten assessment policy. Of the 19 states, 12 allowed localities to determine the assessment measure used, while seven states required the use of state-developed or state-adopted commercial assessments. Required KEAs included Dynamic Indicators of Basic Early Literacy Skills (DIBELS) (required by five states), Phonological Awareness Literacy Screening (PALS) (two states), and Brigance (two states) (Connors-Tadros, 2014). In 2016, the number of states rose to 25 with Teaching Strategies GOLD® and the Work Sampling System (WSS) also implemented as KEAs.

The Center for Enhancing Early Learning Outcomes found that the individual state context and regulations determine whether the KEA was mandatory, as well as the goals set for the KEA use. For states implementing KEAs, three primary purposes emerged: (1) understand and report school readiness at kindergarten entry; (2) improve instruction and practice; and (3) support the transition of children into kindergarten (Weisenfeld, 2017). As late as October 2016, states have continued to struggle with selecting the most effective assessment measure to determine school readiness at kindergarten entry (Weisenfeld; Hanover, 2013; Niemeyer & Scott-Little, 2002). While some states still do not conduct kindergarten entry assessments, some have implemented off-the-shelf instruments, some have adopted a subset of items within an off-the-shelf instrument, while some states have developed their own comprehensive system (Hanover).

State Context.

The state of Florida, which is the state in which the present research was conducted, has had nearly a two-decade history of public schools conducting kindergarten entry assessments to

determine school readiness and provide teachers with instructional information on incoming students. Prior to 1999, school districts screened their kindergarten students with self-selected, district-adopted measures. In 1999-2000, legislation took effect requiring uniform system of determining school readiness for all public-school kindergarten students (FLOEL, 2014). Section 411.01(4)(o), Florida Statutes (F.S.) made provisions for school districts to complete a checklist of 17 school readiness expectations “for each child entering kindergarten, using the results to determine whether or not the child was ready for kindergarten” (FLOEL, 2016, p. 1). From 2002-2003 through 2005-2006, Section 1008.21, F.S., created the School Readiness Universal Screening System (SRUSS) to provide systematic and objective results based on the 17 indicators highlighted in s 411.01(4)(o), F.S. This statute required that each school district complete the Early Screening Inventory – Kindergarten (ESI-K) on all kindergarten students and select one of two observational measures: The Work Sampling System™ or the Ready-for-School Behavioral screener. In 2004-2005, the option between observational measures was removed and all children were assessed with the ESI-K and Dynamic Indicators of Basic Early Literacy Skills™ was added.

In 2006-2007, Section 1002.69, Florida Statutes, was created to implement the Voluntary Prekindergarten (VPK) Education program as well as outline a statewide mandatory kindergarten screening to determine the percentage of children ready to start the traditional K-12 educational system (FLDOE, 2013). This statute repealed s. 1008.21, F.S. and required school districts within the first 30 student contact days to gather information on the children’s readiness for kindergarten based on “performance standards adopted for use in Voluntary Prekindergarten (VPK), currently the *Florida Early Learning and Developmental Standards for Four-Year-Olds (2011)*” (FLOEL, 2016, p. 1). The Florida Kindergarten Readiness Screener (FLKRS) consisted

of a subset of the Early Childhood Observation System™ (ECHOS) and the first two measures of the DIBELS – letter naming fluency and initial sound fluency. Over the next decade, FLKRS measures have been removed and replaced with similar measures of basic emergent literacy skills (i.e., Kindergarten Florida Assessments in Reading Instruction [FAIR-K] replacing DIBELS and WSS replacing ECHOS).

Since 2015-2016, Florida’s kindergarten entry assessment, Florida Kindergarten Readiness Screener (FLKRS) has been comprised solely of an expanded subset of the Work Sampling System™ (WSS). Kindergarten readiness is determined by observational data gathered by kindergarten teachers within the first 30 student contact days. The WSS provides teachers, parents and policy makers with a categorical rating of *Not Yet, In Process, and Proficient* (FLDOE, 2015). The Florida Department of Education provided an operational definition of kindergarten readiness based on the WSS within the 2015-2016 Florida Kindergarten Readiness Screener Administration Manual (FLDOE, 2015) with a rating of *Not Yet* indicating non-readiness for kindergarten with ratings of *In Process* and *Proficient* indicating kindergarten readiness. These ratings are described in detail in Chapter Three.

Statement of the Problem

According to the Organization for Economic and Community Development (OECD), more than three-quarters of all 4-year-olds in the United States are enrolled in some form of early childhood educational program. This percentage of young children who are participating in some type of preschool experience has risen in the past few years. The OECD reports that during the 1990s, enrollment rates of 4-year-olds remained steady at 60%, increased to 65% in 2005 with current enrollment dramatically increasing to 78%. Within the past decade, early childhood education has quickly emerged on the political scene. This spotlight became

highlighted with drastic cuts to Head Start, the federally funded early childhood program, during the government sequestration in the fall of 2013, as well as the spring of 2014 when President Obama announced in his State of the Union address his intention to call upon Congress to expand access to high-quality preschool to every child in America (Obama, 2014). As part of that effort, the President offered the *Preschool for All* initiative as a means to partner with all states to provide all low- and moderate-income four-year olds with high-quality early childhood experiences, while encouraging states to serve four-year-olds from middle-class families. With the president's focus on quality early childhood experiences for all four-year-old children, a deeper level of interest and need for evaluation of school readiness quality of established city- and state-level universal prekindergarten programs and federal programs such as Head Start has emerged. With the increased level of interest in early childhood education there has been a corresponding commitment amongst early childhood researchers/practitioners and federal- and state-level policy makers to ensure that ongoing validation work occurs with the measures used to assess school readiness at kindergarten entry.

Validity of Kindergarten Entry Assessments.

State Education Agencies and policy makers find the selection of an effective assessment tool to measure school readiness at kindergarten entry a challenge. As states have different priorities for early childhood education, by design they select a plethora of methods for developing and implementing kindergarten entry assessments (Hanover, 2013). A current trend of states awarded Race to the Top – Early Learning grants has been to adopt a proprietary off-the-shelf assessment device as a foundation and then customizing to the system to meet the state's standards and priorities (Hanover, 2013). The Work Sampling System™ and Teaching Strategies GOLD® are the most common off-the-shelf assessment measures used as kindergarten

entry assessments, either in the measure’s entirety or adapted versions of the measure (Wat, Bruner, Hanus & Scott-Little, 2012). Hanover (p. 13) found that states that used adapted versions of off-the-shelf measures “pick and choose certain assessment items to use as part of their systems, and discard or de-emphasize remaining items. The retained items are a result of alignment analyses, which identify the assessment indicators most pertinent to state learning standards.”

Although validation studies have been completed on the Work Sampling System™ (WSS) and predictive validity of first grade WSS with 3rd grade state assessments, there are no published studies of the enhanced version of the WSS being used as Florida’s kindergarten entry assessment (Gallant, 2009; Meisels, Jablon, Dichtelmiller, Marsden, & Dorfman, 2001). According to the *Standards for Educational and Psychological Testing* (2014) validation of measures is an ongoing process of gathering relevant evidence to provide sound basis for the proposed score interpretations (p. 11). The authors’ additionally argue “when test scores are interpreted in more than one way (both to describe a test taker’s current level of the attribute being measure and to make a prediction about a future outcome, each intended interpretation must be validated” (AERA, APA, NCME, 2014, p. 11).

Validation can be viewed as a process of constructing and evaluating arguments for and against the intended interpretation of the test scores and their relevance to the proposed use of the measures (AERA, APA, NCME). The authors remind stakeholders that validation is a joint responsibility between test developer and test user. The test user is responsible for evaluating the validity evidence in the specific setting in which the test is to be used. Although studies were conducted validating VPK Assessments with ECHOS and DIBELS, there have been no

published validation studies of VPK Assessments with WSS nor with children within classrooms in a blended Head Start/VPK program.

In a similar fashion, multiple validation studies have been conducted with Teaching Strategies GOLD® that highlight the measure as a valid measure (Heroman, Burts, Berke, & Bickart, 2010; Kim, Lambert, & Burts, 2013; Lambert, Kim, & Burts, 2015). However, there are no published validation studies of using the Teaching Strategies GOLD® to determine school readiness at kindergarten entry using the WSS as the school readiness determinant assessment measure.

Purpose of the Study

The purpose of this quantitative correlational study was to understand how a standards-based, general outcomes measure, Florida VPK Assessments (FLDOE, 2011) (VPK Assessments) and a curriculum-embedded measure, Teaching Strategies GOLD®, collected in the preschool year within a blended Head Start/VPK program related to kindergarten readiness within the state of Florida. The study focused on determining if predictive relationships existed between the independent variables (i.e., VPK Assessments and Teaching Strategies GOLD®) and the dependent variable of school readiness at kindergarten entry using data from the Florida Kindergarten Readiness Screener (FLKRS) Work Sampling System™ (WSS). The dependent variable of school readiness was determined from the WSS categorical results of *Not Ready*, *In Process*, and *Proficient*.

Additionally, the study examined whether the selected child characteristics of gender and ethnicity moderated the predictive relationships between the independent variables of the VPK Assessments and Teaching Strategies GOLD® measures and the dependent variable of school

readiness at kindergarten entry of *Not Ready*, *In Process*, and *Proficient*. Many studies trying to determine if group differences exist often investigate potential differences within gender. The purpose of examining potential group differences within this study based on ethnicity, specifically between Hispanic and non-Hispanic children was due to the demographic distribution of the school district. The school district itself was mainly a homogeneous population when examining race with over 61% of students being white, not Hispanic. However, the fastest growing racial demographic of the school district was Hispanic students with over 23% of the population. As the Hispanic population was growing at a rapid rate within the school district, the decision was made to examine group differences between ethnicity as the variable specifically focused on the determinant of Hispanic or non-Hispanic. From a measurement standpoint, the interactions addressed the question of whether the predictive validity of the VPK Assessments and Teaching Strategies GOLD® were the same for different groups; did the assessments predict school readiness equally for males and females or Hispanic and non-Hispanic children? By understanding the relationships between VPK Assessments and Teaching Strategies GOLD® scores in preschool and levels of kindergarten readiness, policy makers can provide guidelines for assessment practices in early childhood education used for program planning and individualization of learning experiences for young children.

Archival data from two cohorts of blended Head Start/VPK children enrolled in 2014-2015 and 2015-2016 were used for all analyses. As no published studies have been completed related to the predictive validity of the state-mandated VPK Assessments and federally-recommended Teaching Strategies GOLD® for school readiness at kindergarten entry, results from this study may increase the body of knowledge and research related to kindergarten entry

assessments and assessments used in early childhood settings to not only predict school readiness, but to help drive instruction within the early childhood educational setting.

Research Questions

This study was quantitative, longitudinal and non-experimental in nature based on the selected research questions and convenience sample identified for the study. The study used existing data (secondary analysis) from preschool children enrolled in a blended Head Start/state-funded voluntary prekindergarten program (VPK).

This quantitative study addressed the following research questions:

RQ1. What is the relationship between the VPK Assessment subscales (Print Knowledge, Phonological Awareness, Mathematical Concepts, Oral Language/Vocabulary) and Teaching Strategies GOLD® (Social-Emotional, Cognitive, Literacy and Mathematics) and kindergarten readiness (Not Ready, In Process, and Proficient)?

RQ2. To what extent are the relationships between the VPK Assessment subscales (Print Knowledge, Phonological Awareness, Mathematical Concepts, and Oral Language/Vocabulary) and kindergarten readiness (Not Ready, In Process, and Proficient) moderated by gender (female vs. male)?

RQ3. To what extent are the relationships between VPK Assessment subscales (Print Knowledge, Phonological Awareness, Mathematical Concepts, and Oral Language/Vocabulary) and kindergarten readiness (Not Ready, In Process, and Proficient) moderated by ethnicity (Hispanic vs. Non-Hispanic)?

RQ4. To what extent are the relationships between Teaching Strategies GOLD® subscales (Social-Emotional, Cognitive, Literacy, and Mathematics) and kindergarten readiness (Not Ready, In Process, and Proficient) moderated by gender (female vs. male)?

RQ5. To what extent are the relationships between Teaching Strategies GOLD® subscales (Social-Emotional, Cognitive, Literacy, and Mathematics) and kindergarten readiness (Not Ready, In Process, and Proficient) moderated by ethnicity (Hispanic vs. Non-Hispanic)?

Limitations of the Study

As this study was a secondary analysis, as the researcher, I was restricted to the variables that were collected by the state of Florida. Additionally, as I did not collect the data, I needed to rely on the descriptions of the procedures provided that were used to collect and maintain the quality of the data.

This study was also limited by the impact of teachers' self-reporting the Teaching Strategies GOLD® levels of achievement on the authentic assessment. Teachers were trained to collect anecdotal notes, photos and videos that documented the level of achievement each child attained throughout the assessment window. Although teachers were required to participate in mandated online formative training modules and follow-up professional development provided by the school board, ultimately the inter-rater reliability of assessment levels by domain may have differed by teacher perceptions.

Delimitations of the Study

This study was delimited to a blended Head Start/VPK program within the state of Florida. Each program that accepts VPK funding from the state of Florida signs a statement indicating required participation in Assessment Periods one (August and September) and three (May) of the Florida VPK Assessment (FLDOE, 2011) and Teaching Strategies GOLD®

assessment measures. Children within the sample were the age of four years on or before September 1st of the program year. Additionally, Head Start criteria requires that children be enrolled based on the highest need, generally by the income level of the entire family. This eligibility criterion was determined by the United States Department of Health and Human Services guidelines for enrollment year. Priority for enrollment is given to families that are at or below 100% of the poverty guidelines. The gender and ethnicity/race of the sample are presented in Chapter Three.

Definition of Terms

Early Childhood Education (ECE). A branch of education theory comprised of birth through eight years. Within the context of this study, ECE was operationally defined as preschool age children that at the time of enrollment were four-years of age on or before September 1st of the enrollment year. For cohort one, a member of the sample would have turned four-years of age on or before September 1, 2014 and for cohort two, turned four-years of age on or before September 1, 2015.

Blended Head Start/VPK Program. An ECE program that provides services to children from low socio-economic families who have met the Head Start eligibility criterion and were considered four-year of age for the program year. This program is considered blended as the Head Start grant provides resources for 3.5 hours of an academic day, while Florida's VPK funding also provides for 3.5 hours of educational experiences within an academic day. Therefore, the program provides 7.0 hours of educational experiences within a blended funding model that meets Head Start federal standards and the State of Florida VPK standards.

School Readiness at Kindergarten Entry. The state of Florida has defined kindergarten readiness within Section 1002.69, Florida Statutes, (F.S.) as a mandatory statewide

kindergarten screening to be administered within the first 30 days of Kindergarten for all public school children. This Kindergarten screening is known as The Florida Kindergarten Readiness Screener (FLKRS). For the assessment period selected for this study, the FLKRS is comprised of an expanded version of the Work Sampling System™ (WSS). In order to be considered ready to begin Kindergarten, a child must have a WSS composite score of *In Process* or *Proficient*.

Validity. The degree to which evidence and theory support the interpretations of test scores for proposed uses of tests (AERA, APA, NCME, 2014, p. 11).

Importance of the Study

As determining school readiness at kindergarten entry through the use of assessments becomes more universal across the United States, determining if different assessment measures in ECE programs predict readiness has important implications for a variety of stakeholders. Within the state of Florida, state education agencies and policy makers are interested in determining if the state-mandated VPK Assessment and Teaching Strategies GOLD® predict readiness to start school for children participating in ECE programs. Standard 7.12 of the *Standards for Educational and Psychological Testing* (2014, p. 129) states: “When test scores are used to make predictions about future behavior, the evidence supporting those predictions should be provided to the test user.” Information gathered from such studies could provide decision-makers with impartial data to determine if the selected assessments for VPK can predict school readiness and provide parents, teachers, school districts, private providers and policy makers with enough valid data to adjust instruction and supports to help children increase the levels of school readiness as determined on the state-selected kindergarten entry assessment, Work Sampling System™ (WSS).

Organization of the Study

Chapter One introduces the study. It includes the statement of the problem, purpose of the study and the research questions, definitions of the terms used within the narrative, the importance of the study, and the organization of the study. Chapter Two consists of a review of the literature that includes a discussion of early childhood education impact and models, kindergarten readiness assessment policies in the United States, and in Florida; assessments for preschool achievement; and predictive validity for kindergarten readiness.

Chapter Three is the methods section, which describes the participants, the process of data collection, the statistical analyses that were performed, and a summary of the chapter. Chapter Four is the results section, which describes the data collection, descriptive statistical analyses, and results of the main statistical models analyses for each cohort used to answer the research questions. Chapter Five is the discussion section, which provides the findings of the study, limitations of the study, and implications for future research and policy.

CHAPTER TWO

REVIEW OF LITERATURE

The purpose of this chapter is to provide a framework for evaluating the predictive validity of the state-mandated Florida Voluntary Prekindergarten (VPK) Assessment standards-based direct assessment and the curriculum-embedded, observational-based Teaching Strategies GOLD®. The outcome that was used in these predictive validity analyses was the state-mandated kindergarten entry assessment, The Work Sampling System™ (WSS). Chapter Two consists of a review of the literature that includes a discussion of early childhood education impact and models, kindergarten readiness assessment policies in the United States, and in Florida; assessments for preschool achievement; and predictive validity for kindergarten readiness.

Federal and state statutes have placed an emphasis on children enrolled in early childhood educational programs exhibiting specific school readiness domain-specific skills at kindergarten entry. This chapter will establish the formation of an early childhood model with blended federal and state funding and the selection of early childhood assessments, especially why they are used and how they should be used during the preschool year. Additionally, the psychometric concepts of validity and specifically predictive validity will be discussed with a particular emphasis on why these concepts of validity need to be evaluated for the early childhood assessments and kindergarten entry assessment in the state of Florida, *Florida Kindergarten Readiness Screener (FLKRS)* comprised of the WSS.

Early Childhood Education

Overview. Support for our youngest children to have access to a variety of high-quality early childhood educational experiences continues to grow within the United States. One could surmise that the one notion that a majority of adults within the United States -- politicians, researchers, educators and parents -- have agreed upon is that phenomenal things can occur when children are small; and investing in preschool or formalized early childhood education may be no exception. With the dawn of the 21st century, access to quality early childhood educational experiences became a topic of widespread discussion and debate, not only within the educational and research community, but also with national and local politicians. This debate grew out of discussions related to the state of education and birth of K-12 educational accountability with the No Child Left Behind Act of 2001 (No Child Left Behind [NCLB], 2002).

Spurring the debate and eventual action was President George W. Bush's 2002 State of Union Address that focused on his next step of education reform, which was to highlight the need to prepare children to read and succeed in school with improved Head Start and early childhood development programs. Head Start was a federally funded program that began in the summer of 1965 as part of President Lyndon Johnson's War on Poverty. The tenet of Head Start has been to provide and promote school readiness for children in low socio-economic families by offering educational, nutritional, health, social, and other services. Since the inception of Head Start in 1965, more than 32 million children, ages birth to five years and their families, have been served. In 2014, Head Start was funded to serve nearly one million children and expectant mothers within centers, family homes, and in family child care homes located in urban, suburban, and rural communities throughout the nation (ECLKC, 2015a). Within the state of Florida in 2014, the federal government allocated over 318 million dollars to provide enrollment

opportunities to nearly 38,000 children and expectant mothers in a host of settings (ECLKCa). More than 43% of these Head Start opportunities were provided to children within some form of educational center for more than 6 hours a day. In order to provide this full-day model of more than 6 hours daily, most Head Start grantees supplement Head Start funding with a blended funding model with state funding.

When comparing four-year old children who attended a formalized early childhood educational experience versus those who did not, the immediate and long-term effects become compelling with those children exhibiting broader vocabulary and reading skills, improved attendance and higher graduation rates. Current research of children within communities that attended formalized early childhood programs, regardless of family composition or socio-economic status, gained critical social and persistence skills before entering kindergarten that not only heightened their social and academic skills within the traditional K-12 setting, but also their future adults lives within their said communities (Urban Child Institute, 2015). Additionally, children who enter kindergarten exhibiting emergent academic and social skills performed at higher levels through school and completed more academic years of education. In turn, a more educated work force positively correlated with these individuals receiving higher incomes, more public revenues and decreased poverty and crime (Urban Child Institute, 2012).

In 2015, the National Institute of Early Education Research (NIEER) released the annual State of Pre-k report for the 2013-2014 school year. After the devastating impact the federal recession incurred upon all areas of education, the 2013-2014 school year offered hope of a recovery for state-funded pre-K. Although state-funded programs had not fully recovered from the staggering impacts of half a billion dollars in cuts in the 2011-2012 school year, state funding for pre-K increased by nearly \$120 million and served 1,347,272 children in 2013-2014. State

pre-K continued to be largely a program for four-year-old children, with said population accounting for more than 1.1 million, or about 86%, of the pre-K children enrolled (Barnett, Carolan, Squires, Clarke Brown, & Horowitz, 2015).

In addition to the national perspective of the current state of pre-K, NIEER also released state-specific reports. Florida's Voluntary Prekindergarten Education Program (VPK) was created as a result of a 2002 state constitutional amendment requiring access to a high-quality prekindergarten experience for all of Florida's four-year-old children. Parents of age-eligible children were able to choose from two VPK program models: either a 300-hour summer program, which each school district was mandated to offer or a school year program totaling 540 instructional hours. A variety of educational settings provided VPK within the state of Florida, such as licensed child care centers, accredited faith-based centers, licensed family child care homes, public schools, and accredited non-public schools. Since the inception of Florida's VPK program within private centers and school districts in 2005, enrollment has increased from 47% to 80% age-eligible four-year-olds ($N = 170,266$) participating in 2014 (Barnett, Carolan, Squires, Clarke Brown, & Horowitz, 2015). It is interesting to note that within the 2013-2014 sample of age-eligible four-year olds, 80% participated with solely VPK state funding, 9% were participating in a blended funding model with Head Start, 1% were participating in a blended funding model with Special Education, and 10% of age-eligible children participated in a self-contained Special Education model with no additional VPK funding (Barnett, Carolan, Squires, Clarke Brown, & Horowitz).

Early History. Since the 1960s, preschool education has emerged from an experimental idea born from researchers and scholars into a system that is widely accepted by leaders in the worlds of policy, economics, and business. Over the past two decades, the notion of preschool

education as a system that just targets children within the year prior to kindergarten enrollment has morphed into that of an early childhood education system. Multiple funding sources at the state and federal level, as well as hosting a myriad of children on a continuum of universal services based on age to income eligibility for enrollment have comprised the early childhood system. Early childhood education has been widely viewed as perhaps the best means for improving the educational and later-life outcomes of young children by addressing racial and class gaps in educational achievement and protecting the societal investments in education. Over time, the interested parties in early childhood education grew from a handful of scholars to leaders across a variety of fields, including business leaders, philanthropists, advocates, economists, lawyers, and public officials (Zigler, Gilliam, & Barnett, 2011). Wide-ranging interest of early childhood programs from social media to federal budget debates exemplifies the priority of establishing high-quality programs for our youngest and most vulnerable learners and members of our society.

Although the case for establishing early childhood programs has been fervent, there remains a continued practical debate related to program implementation. Years of data collection within the field of early childhood indicated that the return on investment was the highest for disadvantaged children who do not receive abundant amounts of parental investment in the early years (Heckman, 2011). Programs in the 1960s and early 1970s such as the High Scope Perry Project, Carolina *Abecedarian*, and Head Start all focused on intensive preschool programs for socially and economically disadvantaged children. In 2001, Susan Urahn of the Pew Charitable Trusts examined the data on early childhood and found that it could have profound impact on children's school and life success. However, Urahn learned that despite decades of hard work by advocates, both foundation funding and policy makers' interest had not yet caught up to the

research evidence on the benefits of high-quality early education (Watson, 2011). Based on Pew's criteria for selecting issues for building policy campaign's, Urahn determined that the early childhood movement encompassed several factors indicating 'ripeness' of the potential policy: the growing body of research showing the importance of early brain development (Phillips & Shonkoff, 2000); decades of advocacy for child care funding; interest within funders in the field; and the increasing understanding that many of the building blocks of educational success (or failure) were built in the earliest years. Urahn's considerations also were informed by the experience of the three states (Georgia, New York and Oklahoma) that had already made commitments to providing early childhood experiences for all four-year-olds. At this time, policy makers and parents did not understand or recognize early childhood education as a powerful reform rather than just the two years of childcare prior to entering kindergarten (Watson). After analyzing and appraising the importance and potential impact of early childhood education, Pew designed and communicated the *Pre-K Now* initiative in September 2001. Pew's *Pre-k Now* campaign and the Bush Administration's early childhood initiative *Good Start, Grow Smart* (2002) gave rise to the foundations of universal early childhood programs (Watson).

More recently, Obama made instilling and providing quality early childhood educational experiences for all children a fundamental priority of his second term of office. The president first proposed the *Early Learning Initiative* in early 2013. The president's proposal included a series of new investments that focused on providing a continuum of high-quality early learning services for all children, not only those from lower socio-economic backgrounds. The *Early Learning Initiative* called for a support system ranging from birth through age five years that included providing Preschool for All, instituting Early Head Start-Child Care partnerships and

expanding home visiting programs (ECLKC, 2015b). Reference Appendix A for a detailed historical description of the emergence and growth of the Head Start/Early Head Start program.

Emergence of Universal Prekindergarten

Initially, the effects of the No Child Left Behind (NCLB, 2002) and Race-to-the-Top (U.S. DOE, 2009) legislation and accountability began to be felt in early childhood education as policy makers believed that an early start on developing academic skills would help children reach the standards they were expected to achieve in elementary school (Stipek, 2006). NCLB exerted a heavy influence on states and focused their attention on defining expectations for what young children should learn before entering kindergarten in an effort to meet accountability targets in the later elementary grades continuing on towards graduation rates. In 2015, a vast number of governors discussed within their State of State addresses that early learning experiences were at the center of the states' strategies to reduce the ever-increasing achievement gap and ensure long-term student success. Eleven governors promised funding to develop or expand early childhood education and kindergarten opportunities, especially for families from a low-socio-economic background (Aragon & Rowland, 2015). Since the beginning of the 2014 fiscal year, four states highlighted first-time investments in early childhood programs for their 3- and 4-year olds with total state funding for early childhood programs increasing by twelve percent nationwide. Aragon and Rowland (2015) found at the beginning for the 2015 fiscal year, only 6 states did not provide funding for preschool. As the universal preschool education movement continued to grow and President Obama proposed programs to promote high-quality early childhood education for all, the debate related to the need for standardization of educational opportunities and quality of services.

Universal Prekindergarten in Florida. In November of 2002, Florida voters approved a state constitutional amendment that no later than the beginning of the 2005 school year, each four-year-old child within the state would have access to a prekindergarten opportunity in the form of an early childhood development and education program. The program was to be voluntary for parents, high quality, free of cost and delivered according to professionally accepted standards (FLOEL, 2011). As stated in Section 1002.53(1), Florida Statutes, the Voluntary Prekindergarten Education (VPK) program was available for any child who was a Florida resident and attained four years of age on or before September 1 of the academic year. Parents of these eligible children who accepted the opportunity to participate in the VPK program could choose either a 540-hour school year or 300-hour summer program offered by either a private prekindergarten provider or public school. Within the VPK statute language was verbiage that each of the 67 school districts across the state would be required to provide at least one session of a 300-hour summer program to provide services to children who had not used their voucher for the school-year program. This program would be geographically located to provide children who did not utilize the 540-school year program with a foundational experience before entering kindergarten.

In 2004, the Florida Legislature enacted this VPK legislation in accordance with the Florida Constitution. Within the language of the legislation, responsibilities for the daily management of the program was given to the Office of Early Learning (OEL) within the Agency of Workforce Innovation (AWI); licensing and credentialing to the Department of Children and Families (DCF); and the creation of standards, curriculum and accountability to the Department of Education (DOE). However, effective July 1, 2011, and allowing for a 3-month transition period ending October 1, 2011, Chapter 2011-142, Laws of Florida, transferred the housing of

the Office of Early Learning from within AWI to DOE (FLOEL, 2011). These agencies were charged to work together to provide leadership and support to local early learning coalitions, school districts and public and private providers to ensure the appropriate implementation of universal voluntary prekindergarten services to Florida's four-year-olds.

Early Childhood Learning Standards and Assessments

Overview. In order to create a standardization of educational opportunities and experiences, a variety of individuals within the early childhood research and policy making communities found that a barrier to overall standardization was a lack of consensus on the operational definition of the construct *School Readiness* (Ackerman & Barnett, 2005; American Federation of Teachers, 2003; Committee for Economic Development, 2002; Kolker, Osborne, & Schnurer, 2004; Mashburn & Henry, 2004; Maxwell & Clifford, 2004; Mead, 2004; National Governor's Association, 2005; Snow, 2006). Additionally, Snow (2006) stated that the lack of an operational definition is also compounded by the vast range of measures implemented within the evaluation of state-funded early childhood programs, as well as the relative lack of agreement within states as to which measure to utilize to either drive instruction or for accountability purposes.

Continuity between early childhood education and primary grades experiences has long been considered essential to children's development. Thus, the relationship between early learning standards for early childhood and K-12 standards was a critical element and topic of discourse to which all states attended. However, there were varying degrees of linkage ranging from early learning standards that were actually incorporated into the K-12 standards to two individual stand-alone sets of standards that simply were created to use similar formats. States indicated that the standards were designed to apply to all children, including children from

diverse backgrounds and children with disabilities; however, conspicuous by their absences were the domains of approaches towards learning and social and emotional development. Kagan and Scott-Little (2004) noted that less emphasis on these areas was due to the fact that states had a tendency to align early learning standards with the K-12 education, which generally focused less intensely on these domains. Kagan and Scott-Little theorized that these domains may have been given less attention because some find that they are often more difficult to operationalize. However, these domains are inextricably related to children's later development and success in school.

Kagan and Scott-Little (2004) clearly articulated this dichotomy between a desire for quality school readiness opportunities while acknowledging a deep-seated and long-standing skepticism toward learning standards among the early childhood learning community. Years of scholarship have indicated that young children's development is anything but standardized. Development during the preschool years unfolds unevenly, is highly individualized, and is characterized by growth spurts, developmental lags and irregularity across developmental domains. Although normal and expected, such variation made the argument for the development of standards and assessment not only difficult, but some argued, unwise. The National Education Goals panel (NEGP) identified five domains: (1) physical and motor development; (2) social and emotional development; (3) approaches toward learning; (4) language and literacy; and (5) cognitive and general knowledge, which includes mathematics. In response to A Nation at Risk (United States, 1983) and The National Reading Panel's Report: Teaching Children to Read (US, 2000) the most commonly addressed domain within early childhood state standards has been language and literacy, while approaches to learning was the least commonly included domain, with only seven states addressing this domain.

Good Start, Grow Smart Initiative. The federal government went one step further and defined expected outcomes for Head Start children and the *Good Start, Grow Smart Initiative*. This federal initiative, Good Start Grow Smart, had substantial impact on states' early childhood education initiatives. The initiative called for states to develop early learning standards for children ages 3-5 years in language, literacy and mathematics that were aligned with their Kindergarten through twelfth grade (K-12) standards. Although in 1998, there was language built into the Head Start Reauthorization Act that emphasized the need for academic skills with developmentally appropriate alignment with states' K-12 standards, this trend was only embraced by providers of services to lower socio-economic children. Previous to the advent of the Good Start Grow Smart initiative, fewer than half of the states had preschool standards (Neuman & Roskos, 2005). Although early childhood education scholars and experts agreed that clearly articulated standards could provide direction and coherence to early childhood educational experiences, however, similar concerns expressed about K-12 standards being overly narrow and comprised of a laundry list of skills apply, perhaps even more significantly, to preschool. Critics stated that the fear over these narrow standards in K-12 being pushed down to apply to younger children could be even more developmentally inappropriate and potentially do greater damage to preschool age children (Stipek, 2006). Although standards were not created to make high-stakes decisions about individual children, the possibility that a child's progress on skills and abilities articulated in the standards might be used to make placement decisions or to determine whether a child would go on to kindergarten remained worrisome.

As early childhood standards began to be shared within the educational landscape, the focus on the development of standards was on three- to five-year-olds or for those who were nearing entry into the public schools. Thus the standards had a greater potential to be used as

“gatekeepers” for kindergarten than as the basis for developmental pedagogy and instruction. It is important to note that few states have instituted safe guards against such a use of standards (Kagan & Scott-Little, 2004).

Within the continued dialogue of establishing standards, early childhood experts were concerned that as standards created after 2001 were often created in the form of laundry lists of skills, this implementation of standards-based instruction would lead to fragmented teaching of isolated skills that may not be meaningful or motivating for young children. Children’s development is strongly affected by what they experience in the world-interactions with their parents, exposure to literacy-rich environments, and experiences in preschool. These experiences varied widely and many in the field of early education question the utility, viability and equity of standard expectations for young children, particularly if said standards will be used in the future to assess the children and perhaps screen them out of services for which they are labeled ‘unready’. There was widespread concern regarding the content of early learning standards. Many in the field feared that said standards would be more academically focused than developmental in orientation; while other early education individuals were concerned that the standards would have favored certain domains that could have the ability to slant the very fiber of early education principles and foundations. Emergent literacy experts were quick to point out that although children need to develop the decoding skills (phonological awareness, alphabetic principles, repertoire of automatically recognized words) that are typically emphasized in standards, children’s language, conceptual and cultural knowledge, vocabulary and verbal reasoning abilities were just as important as decoding to reading success. As discussion related to standards increased, experts reminded the authors of the standards that emergent literacy skills were interdependent of each other and were best learned in the context of meaningful text and

not in isolation in order to complete a checklist of standards of skills (Neuman & Roskos, 2005).

Many early childhood experts agreed that young children also need to learn to make connections between the text they read or hear being read to them and their own experience, and previous texts is vital for growth in emergent literacy skills. Just as it was important to early childhood experts that standards included within emergent literacy were not developed in isolation of other academic and social skills, the same belief system was communicated in relation to cognition and mathematics. If one considered a typical standard for young children (e.g., children can count to 20), counting would only be meaningful to children if they also understand one-to-one correspondence between objects and numbers and that the last number when counting refers to how many items were counted. Early childhood experts continued to communicate that standards that did not articulate these other understandings could lead to mindless instruction, such as rote counting. Early mathematics learning has also been comprised of an understanding of shapes, measurement (such as mass, length and weight), comparisons, ordering, patterns, classifying and organizing and representing objects (Clements, Sarama, & DiBiase, 2004). These varied and interdependent components of mathematics needed to be included in standards and taught in contexts that were meaningful to young children.

Early childhood scholars and experts quickly shared with the authors and interpreters of standards related to young children that although K-12 standards often were void of social-emotional constructs/concepts, these early childhood state standards needed to include the many social and affective dimensions of children's development that affect their academic success and their lives. Studies conducted around the millennium demonstrated the value of a positive social context in classrooms for promoting academic achievement (Hamre & Pianta, 2005). Moreover,

it has been postulated that children's social skills predict their learning as well as affect their relationships with peers and adults in school (Miles & Stipek, 2006). There has also been compelling evidence that emotional adjustment and self-regulation at school entry age are strong predictors of children's academic performance in school, over and above their academic skills at school entry. Early education scholars continually state that if the purpose of early childhood standards were designed to influence the content and focus of preschool programs, it must be imperative that the standards include all aspects of development that research indicates are important for young children's development. To summarize, implicit early childhood standards for student learning can be valuable, but only if the standards are well informed and created by knowledge about how young children learn and the many dimensions of development that contribute to children's success in and out of school (Zigler, Gilliam, & Barnett, 2011).

Snow and Van Hemel (2008) stated that within early childhood assessments, these assessments should accomplish four main purposes. Assessments should: (1) determine an individual child's level of functioning; (2) guide intervention and instruction; (3) evaluate the performance of a program or society; and (4) advance knowledge of child development.

Head Start. In 1999, Congress mandated Head Start implement Child Outcome Standards of Learning in early literacy, language and numeracy skills; these skills had not been fully and effectively implemented by 2001. Examples of the standards of learning goals of children in Head Start were: (a) develop phonemic, print, and numeracy awareness; (b) understand and use increasingly complex and varied language; (c) understand and use language to communicate for various purposes; (d) develop and appreciate an appreciation of books; (e) for non-English background children, progress toward the acquisition of the English language; (f) know that letters of the alphabet are a special category of visual graphics that can be

individually named; (g) recognize a word as a unit of print; (h) identify at least ten letters of the alphabet; and (i) associate sounds with written words.

The Bush administration planned on strengthening Head Start by ensuring that local Head Start programs were evaluated on whether they effectively prepare enrolled children to meet the standards of learning, as well as training teachers to use the most effective, research-based methods of early reading and language skills instruction to better meet the needs of the children. The Office of Health and Human Services was charged with developing and implementing an accountability system by the fall of 2003 to assess each Head Start center on the standards of learning in the areas of early literacy, language, and numeracy. The National Reporting System (NRS) was a standards-based assessment system field tested in the 2002-2003 fiscal year. NRS was a tool that was to be systematically administered one-on-one with children. The data gathered from the NRS were required to be entered into a federal database for scoring and review. For the first time since the inception of Head Start, each local Head Start program was required to assess all children between the ages of three- and five-years old on required learning standards at the beginning, middle and end of each year. In addition, each local Head Start program was to gather and analyze the data throughout the year to determine the progress of the children with a federal monitoring team conducting on-site monitoring reviews related to the implementation of the accountability system.

As previously discussed, the law continued to outline systematic and standardized methods of accountability within Head Start programs with increased federal oversight. However, within the Act was also language that terminated the use of the National Reporting System (NRS) within Head Start programs. Although programs were no longer required to use the NRS tool to assess the progress of children, each program was required to

select a research-based assessment system to continue to collect, analyze and report academic and social competencies and skills of enrolled three- to five-year-old children at the beginning, middle and end of enrollment. Based on the assessment tool programs adopted, the Office of Head Start was receiving data systematically, but with different expectations and outcomes making program oversight and monitoring a continued challenge at both the local and federal level.

In 2007, the 110th Congress approved the *Improving Head Start for School Readiness Act*. This new law terminated the use of the National Reporting System (NRS), however the law increased the level of accountability through strengthened monitoring and oversight, particularly in the implementation of the Child Outcome standards and progress monitoring of children throughout the program year. Head Start Performance Standard §1302.32 (b)(1) states: a Head Start program must:

...conduct standardized and structured assessments, which may be observation-based or direct, for each child that provide ongoing information to evaluate the child's developmental level and progress in outcomes aligned with the Head Start Early Learning Child Outcomes Framework: Ages Birth to Five. (ACF, 2016, p.30)

Additionally, Head Start Performance Standard §1302.32 (c)(1) provides guidance that assessments “must be valid and reliable for the population and purpose for which they will be used, including by being...age, developmentally, culturally and linguistically appropriate and appropriate for children with disabilities” (ACF, 2016, p. 31).

Once a Head Start program has selected an assessment that is valid and reliable, the program must establish and maintain a system for collecting, analyzing and utilizing data for continuous program and child-level improvement. Head Start Performance Standard §1302.102

(c)(2)(i) states the data system must ensure child-level assessment data are aggregated and analyzed at least three times a year, including for sub-groups indicated by the program's community needs assessment as needing additional support, such as dual-language learners, children with disabilities, sex and race (ACF, 2016).

As the 2007 Head Start Act placed an enormous emphasis on using valid and reliable early childhood assessments for children ages three to five years for individual and program accountability purposes, many Head Start programs were ill equipped with the technical skills or time to select appropriate measure(s) (Halle, Zaslow, Wessel, Moodie, & Darling-Churchill, 2011). In order to support Head Start program nationwide, The Office of Planning, Research and Evaluation examined and provided information on six valid and reliable individual child assessment measures for Head Start programs to review prior to selecting and implementing for their specific population. The six individual assessment measures summarized were: (1) Creative Curriculum Developmental Assessment (now Teaching Strategies Gold®); (2) Galileo Preschool Assessment Scales; (3) High Scope Child Observation Record (COR); (4) Learning Accomplishment Profile – 3rd Edition; (5) Mullen Scales of Early Literacy; and (6) Work Sampling System for Head Start (Halle et. al).

In the spring of 2011 after reviewing the assessment measures listed above, the program within this current study selected the Teaching Strategies GOLD® online assessment system (Heroman, Burts, Berke, & Bickart, 2010). Teaching Strategies GOLD® online assessment system is an authentic, observation-based performance assessment system of young children. Authentic assessment systems are ongoing with teachers/caregivers gathering daily anecdotal data within daily routines and settings rather than an additional data collection cycle or source. *Teaching Strategies GOLD®* has 36 objectives organized within the areas of approaches to

learning, language, cognition, literacy, mathematics, and physical and social-emotional development. Teachers gather student anecdotal data within daily activities as the teachers observe and interact with children and families (Heroman, Burts, Berke, & Bickart). Based on these stated functions of Teaching Strategies GOLD®, this assessment addresses two of the four assessment purposes highlighted by Snow and Van Hemel (2008): *determining an individual child's level of functioning*; and *guiding intervention and instruction*.

Voluntary Prekindergarten in Florida. The growing pains of standards being promulgated and overlapping or contradicting each other has not escaped the state landscape. Adhering to federal dispensation, the state of Florida released the Florida Early Learning and Developmental Standards for Four-Year-Olds (2011) for both the school readiness and voluntary preschool (universal VPK for four-year olds) programs. However, this one set of standards for four-year-olds were born out of multiple evolutions of the Florida School Readiness standards.

As the standards movement began to evolve across the nation, Florida's parallel support of early childhood programs expanded with collaboration of public and private early childhood partners (FLDOE, 2011b). In 2000, the School Readiness Act called for the creation of standards for all school readiness programs. In 2000, OEL and DOE promulgated standards for five-year-olds with the original standards being expanded in 2001 to include standards for three- and four-year-olds. In 2004, OEL promulgated standards that were approved from Birth to three-year olds creating a standards-based system ranging from birth through kindergarten entry.

The original Florida Voluntary Prekindergarten Education Standards (2005) were the result of a collaborative review of the Florida School Readiness Performance Standards with

experts in the areas of early children, with an emphasis on emergent literacy and early learning practitioners. This collaborative review concluded that the existing standards were adequate, but needed additional explanatory language and indicators. The original Language and Communication domain was separated into two domains: Language and Communication and Emergent Literacy. These standards were revised and benchmarks were added. In the spring of 2005, the State Board of Education approved the Florida Voluntary Prekindergarten Education Standards (2005) and the Department of Education communicated a commitment to review these standards every three years.

In order to address the commitment to review standards every three years, the Department of Education established a panel of experts to ensure that the 2008 review of the Florida Voluntary Prekindergarten Education Standards (2005) was based on the most current research and evidence-based, effective practices in early childhood education, mathematics and science. In the original Florida Voluntary Prekindergarten Education Standards (2005), the Cognitive Development and General Knowledge domain included four areas: mathematics, scientific thinking, social studies, and the arts. In the summer of 2008, the State Board of Education formally adopted the Florida Voluntary Prekindergarten Education Standards.

In preparation for the forthcoming three-year review of VPK standards in 2011, AWI and DOE began a collaboration to create one set of standards for four-year olds throughout the state, regardless if funded through School Readiness or VPK funding. AWI and DOE formed a state panel of experts, who in an effort to align with a national focus on early childhood standards in five domains of child development, made the collaborative decision to include the domains of: (1) Physical Development; (2) Approaches to Learning; (3) Social and Emotional Development; (4) Language, Communication and Emergent Literacy; and (5) Cognitive Development and

General Knowledge (including Mathematics). The Florida Early Learning and Developmental Standards for Four-Year-Olds (2011) were presented to the State Board for approval as a common framework and language of skills that four-year-old children should be able to exhibit by the end of the prekindergarten year. These standards established one set of performance standards for Florida’s four-year-olds in School Readiness and VPK programs, which were aligned with the kindergarten Next Generation Sunshine State Standards, and Common Core State Standards and which prepared children for kindergarten.

Under the direction and guidance of Dr. Christopher Lonigan, the Florida Center of Reading Research (FCRR) worked in partnership with the Florida Department of Education to create a standards-based assessment system. The VPK Assessment was developed to be able to be used as both a screening and progress monitoring measure. The VPK assessment was created to provide VPK teachers with reliable and valid means of identifying children who were not on a trajectory of success to be “kindergarten ready” in terms of their reading-related and math skills during enrollment in VPK. Once children are identified as being at risk for meeting kindergarten readiness standards, teachers can use the data from the VPK Assessment to provide enriched experiences and focused instructional activities to help children acquire the academic skills to potentially put them on the path to kindergarten readiness.

The VPK Assessment includes progress monitoring measures in the areas of Print Knowledge, Phonological Awareness, Mathematics, and Oral Language/Vocabulary that are aligned with the state standards. Initial item development involved creating sets of items that were aligned to the domains of early literacy and early mathematics that were foundational within Florida’s VPK standards; specifically, items that addressed a range of abilities in oral language, phonological awareness, print knowledge, and mathematics. Additionally, within the

specific domain, items were created utilizing different formats to identify the variety of manners of assessing children's skills (i.e., question stems, response formats) that would be effective with a wide continuum of VPK teachers with limited assessment training. Based on these stated functions of the VPK Assessments, these assessments address three of the four assessment purposes highlighted by Snow and Van Hemel (2008): *determining an individual child's level of functioning; guiding intervention and instruction; and evaluating the performance of a program or society.*

School Readiness at Kindergarten Entry

Over the past decade, federal investments in early childhood assessments have been born out of a national need for developmentally appropriate, psychometrically sound assessment measures to screen and monitor young children, as well as evaluate the overall effectiveness of their early learning programs (Goldstein, McCoach, & Yu, 2018). Kindergarten is a seminal point in children's lives as they move from a variety of early learning and developmental settings and into the traditional K-12 educational system (Goldstein, McCoach, & Yu). As of 2018 in the United States, 35 out of 50 states have either piloted or implemented kindergarten readiness assessments (KEAs) with at least seven states in the development stages (Pierson, 2018). KEAs are designed to be administered by the child's classroom teacher with the data collected being used at not only a baseline for kindergarten instruction and future progress, but as a cumulative snapshot of the child's previous early learning experiences (Goldstein, McCoach, & Yu). As the definitions of the knowledge and skills measured at kindergarten entry varied within states, a subset of states required schools to use specific instruments to address all five essential domains of school readiness (Language and Literacy; Cognitive and General Knowledge [including early mathematics and early science]; Approaches Towards Learning; Physical Well-being and Motor

Skills; and Social-Emotional Development) (U.S. Department of Education, 2011, 2014; Goldstein, McCoach, & Yu, 2018).

Miller-Bains, Russo, Williford, DeCoster and Cottone (2017) reported that a majority of states selected performance-based, authentic observational assessment measures for adoption as the state-wide mandated KEA. K. Snow (2011) highlighted that performance-based assessment measures allow for teachers to rate a child's performance on a set of items after a defined period of observing and documenting the child's abilities and functioning within the naturally occurring classroom environment/learning situations. Early Education practitioners and researchers share that performance-based assessments are more developmentally-appropriate for young children (May & Bagnato, 2010; K. Snow, 2011; Miller-Bains et al., 2017), enable the measurement of skills as behaviors as they naturally occur rather in an artificial or unfamiliar setting (Bagnato & Macy, 2010; Dennis, Rueter, & Simpson, 2013; Miller-Bains et al., 2017) and can pose less imposition on teachers, children, and classroom time as skills are assessed during regular instruction (McAfee & Leong, 2011; K. Snow, 2011; Miller-Bains et al., 2017).

Although the use of state-mandated performance-based KEAs was praised by educational practitioners, researchers reminded stakeholders that the advantages of the assessments must be considered along with the intended use of the ratings of the children's skills on the KEA (Miller-Bains, et al.). It is imperative that the psychometric properties of reliability and validity of the KEA provide understanding of the assessment measures' use for a variety of purposes as well as the extent to which stakeholders can trust the results produced by the measure (C. Snow & Van Hemel, 2008). This is vital for large-scale KEAs as many stakeholders are able to gain insight into the same set of skills for individual students and aggregated to the class, school, district and state-levels (C. Snow & Van Hemel, 2008; K. Snow, 2011).

In 2006-2007, Section 1002.69, Florida Statutes, was created to implement the Voluntary Prekindergarten (VPK) Education program as well as outline a statewide mandatory kindergarten screening to determine the percentage of children ready to start the traditional K-12 educational system. This statute repealed s. 1008.21, F.S. and required school districts to within the first 30 student contact days to gather information on the children’s readiness for kindergarten based on “performance standards adopted for use in Voluntary Prekindergarten (VPK), currently the *Florida Early Learning and Developmental Standards for Four-Year-Olds (2011)*” (FLOEL, 2016, p. 1). The Florida Kindergarten Readiness Screener (FLKRS) consisted of a subset of the Early Childhood Observation System™ (ECHOS) and the first two measures of the DIBELS – letter naming fluency and initial sound fluency. Over the next decade, FLKRS measures have been removed and replaced with similar measures of basic emergent literacy skills (i.e., Kindergarten Florida Assessments in Reading Instruction (FAIR-K) replacing DIBELS and WSS replacing ECHOS).

Since 2015-2016, Florida’s kindergarten entry assessment, Florida Kindergarten Readiness Screener (FLKRS) has been comprised solely of an expanded subset of the Work Sampling System™ (WSS). The Work Sampling System (WSS) (Meisels, Jablon, Dichtelmiller, Marsden, & Dorfman, 2001) is an observational authentic assessment for children from preschool (age 3) through Grade 6. The publisher states that the assessment has been used with more than 850,000 children, mostly in prekindergarten and kindergarten enrolled in nearly 10 states, New York City and most recently, Florida (Meisels, 2011). WSS is a criterion-based, curriculum-embedded performance assessment that was created to document and determine the state of what children are learning and have begun to master by providing information on individual students’ academic and social achievements (Meisels). The measure allows teachers

to collect a wide-span of information from multiple sources and focus the observations to evaluate what children know and can do during the specific assessment window. Customized versions of the WSS have been created by the publisher, NCS Pearson for State Education Agencies (SEAs), Local Education Agencies (LEAs), and Head Start.

Kindergarten readiness is determined by observational data gathered by kindergarten teachers within the first 30 student contact days. The WSS provides teachers, parents and policy makers with a categorical ratings of *Not Yet*, *In Process*, and *Proficient* (FLDOE, 2015). The Florida Department of Education provided an operation definition of kindergarten readiness based on the WSS within the 2015-2016 Florida Kindergarten Readiness Screener Administration Manual (FLDOE, 2015) with a rating of *Not Yet* indicating non-readiness for kindergarten and ratings of *In Process* and *Proficient* indicating kindergarten readiness. Based on these stated functions of the FLKRS WSS, this assessment addresses three of the four assessment purposes highlighted by Snow and Van Hemel (2008): *determining an individual child's level of functioning; guiding intervention and instruction; and evaluating the performance of a program or society.*

Validity

Overview. According to the *Standards for Educational and Psychological Testing* (2014) validation of measures is an ongoing process of gathering relevant evidence to provide sound basis for the proposed score interpretations (p. 11). The authors additionally argue “when test scores are interpreted in more than one way (both to describe a test taker’s current level of the attribute being measure and to make a prediction about a future outcome, each intended interpretation must be validated” (AERA, APA, NCME, 2014, p. 11).

Validation is a process of constructing and evaluating arguments for and against the intended interpretation of the test scores and their relevance to the proposed use of the measures (AERA, APA, NCME). The authors remind stakeholders that validation is a joint responsibility between test developer and test user. The test user is responsible for evaluating the validity evidence in the specific setting in which the test is to be used. Messick (1989) stated that validity may be defined as “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessments” (pp. 13-14). Rathvon (2004) agreed with Messick’s notion of validity that it evolves over time as new data and findings are continuously shared; acquired through “a continuing process of theoretical and empirical analysis” (p. 53).

Predictive Validity. Although predictive validity has been viewed within criterion validity, it focuses on the relation of scores within a criterion of interest (Cronbach & Meehl, 1955). More specifically, predictive validity is established by how well performance on an assessment predicts later performance on a criterion assessment instrument (Rathvon, 2004; Snow & Van Hemel, 2008). Additional researchers have expressed that it is essential to determine strong predictive validity in early childhood assessment due to the important role that these assessments have in predicting future educational and social outcomes, often leading to decisions related to individualized instruction and educational placement (Caffrey, Fuchs, & Fuchs, 2008; Snow & Oh, 2011). Within this study, the predictive validity focus is on the state-mandated VPK Assessments and Teaching Strategies GOLD® in predicting school readiness at kindergarten entry, as measured by the Work Sampling System™.

Present Study

Taking into account the purpose of the current Florida Kindergarten Readiness Screener (FLKRS) which was comprised of the Work Sampling System™ (WSS) assessment to determine school readiness at kindergarten entry, data on the predictive validity of the VPK Assessments and Teaching Strategies GOLD® during the preschool year must be understood in order for the state of Florida to predict school readiness at kindergarten entry. If strong evidence of predictive validity and relationships exist, the data from the assessments or specific subtest will have potential to assist the state and Head Start policy makers and individual programs to provide appropriate understanding of four-year-old children's school readiness domain-specific skill levels, as well as with informing instruction and interventions for young children in academic and social-emotional skill development. This information may help improve children's future literacy, mathematics and social competencies.

After conducting a literature review, few studies were found examining the Florida VPK Assessments. Several studies examining the psychometric properties representing different ethnic, racial, language, functional status and age groups found Teaching Strategies GOLD® to be a psychometrically promising instrument (Lambert, Kim, & Burts, 2015). The authors continued to examine evidence for the reliability and validity of the data provided by Teaching Strategies GOLD® using two national samples ($n_1 = 10,963$, $n_2 = 20,970$). Based on the results of the study, the researchers found support for the reliability and validity of the scores based on teacher ratings, including confirmatory factor analysis, classical and modern indexes of reliability, and inter-rater reliability statistics (Lambert, Kim & Burts).

Although the literature review did not reveal any studies exploring the relationships between Teaching Strategies GOLD® and the Florida VPK Assessments (a standards-based,

direct assessment), a concurrent validity study was conducted using a sample of three- and four-year-old child ($n = 1,241$) examining Teaching Strategies GOLD® with the Bracken School Readiness Assessment (Panter & Bracken, 2009). The Bracken is an individually administered, norm-referenced assessment of a child's progress in the areas of color, letters, numbers/counting, size/comparison, and shapes. Lambert, Kim and Burts (2015) found that accounting for teacher ratings and clustering effects, moderate relationships were found between Teaching Strategies GOLD and the Bracken School Readiness Assessment.

As the state of the Florida selected the Work Sampling System™ (WSS) for the kindergarten entry assessment, it was important to determine the validity of this assessment. The Pittsburgh Work Sampling Achievement Validation Study was conducted in 1996-1997 with a cross-sectional sample of students enrolled in kindergarten through third grade nested within 17 classrooms (Meisels, Bickel, Nicholson, Xue, & Atkins-Burnett, 2001). The teachers collected data using the Work Sampling System™ (WSS) and administered the Woodcock-Johnson, Revised (WJ-R) in the fall and spring to measure achievement. A regression analyses was completed to determine if the Work Sampling System™ (WSS) made a unique contribution to the child's performance on the (WJ-R). The researchers examined the concurrent aspects of the validity of the WSS, specifically how the teacher ratings correlated with the students' standardized achievement scores on the WJ-R. Meisels et al. (2001) stated that over three fourths of the correlations were within the range of .50 – .75. Additionally, 48 of the 52 correlations between the WSS and the comprehensive scores (broad reading, broad writing, language and literacy, and broad math) fell within the moderate to high range. These results indicated that the Work Sampling System™ (WSS) correlated well with a standardized

assessment and was a reliable predictor of achievement ratings in kindergarten through third grade.

As of January 2017, the Center on Enhancing Early Learning Outcomes (CEELO) reviewed the National Institute for Early Education Research (NIEER) annual state of preschool yearbook to determine the number of states implementing kindergarten entry assessments and the measures utilized. Of the 29 individual states and three state consortia, nine states reported selecting and implementing Teaching Strategies GOLD®, while only one state, Florida, indicated the Work Sampling System™ (WSS) was the kindergarten entry assessment (Weisenfeld, 2017).

Within the state of Florida, children within classrooms in schools with blended Head Start/VPK funding are mandatorily exposed to the standards-based VPK Assessment in the fall and spring and a program-selected assessment recommended by the Office of Head Start at least three times a year. In order to align with Head Start Performance Standards, these assessment measures are often authentic in nature and more curriculum-embedded. The children within these samples were assessment with the Teaching Strategies GOLD® four times a year (September, January, March and May). All students were then observed and measured with the Work Sampling System™ (WSS) as the kindergarten entry assessment. This aim of this study was to examine the predictive validity of the VPK Assessments and Teaching Strategies GOLD® during the preschool year in predicting school readiness at kindergarten entry.

CHAPTER THREE METHOD

In this chapter I review the purpose of the study and the research questions. I then describe the research design and data sources for this secondary analysis. As part of the description of the data sources, I provide information about the two major measures examined in this study, the Florida VPK Assessments and Teaching Strategies GOLD®, as well as the criterion variable of school readiness at kindergarten entry, as measured by the Work Sampling System™ (WSS). Lastly, I provide details about the procedures, participants, and data analyses.

Purpose of the Study

The purpose of this quantitative correlational study was to understand the predictive validity of the Florida VPK Assessments and Teaching Strategies GOLD for children in classrooms within schools in a blended Head Start/VPK program in Florida. The study focused on determining if predictive relationships exist between the independent variables (i.e., VPK Assessments and Teaching Strategies GOLD®) and the dependent variable of kindergarten readiness. Additionally, the study examined whether selected child characteristics of gender and ethnicity moderated the predictive relationships between the independent variables (i.e., VPK Assessments and Teaching Strategies GOLD®) and the dependent variable of kindergarten readiness. By understanding the predictive relationships between VPK Assessments and Teaching Strategies GOLD® scores in preschool and levels of kindergarten readiness, policy makers can provide guidelines for assessment practices in early childhood education used for program planning and individualization of learning experiences for young children.

Research Questions

This quantitative study addressed the following research questions:

RQ1. What is the relationship between the VPK Assessment subscales (Print Knowledge, Phonological Awareness, Mathematical Concepts, Oral Language/Vocabulary) and Teaching Strategies GOLD® (Social-Emotional, Cognitive, Literacy and Mathematics) and kindergarten readiness (Not Ready, In Process, and Proficient)?

RQ2. To what extent are the relationships between the VPK Assessment subscales (Print Knowledge, Phonological Awareness, Mathematical Concepts, and Oral Language/Vocabulary) and kindergarten readiness (Not Ready, In Process, and Proficient) moderated by gender (female vs. male)?

RQ3. To what extent are the relationships between VPK Assessment subscales (Print Knowledge, Phonological Awareness, Mathematical Concepts, and Oral Language/Vocabulary) and kindergarten readiness (Not Ready, In Process, and Proficient) moderated by ethnicity (Hispanic vs. Non-Hispanic)?

RQ4. To what extent are the relationships between Teaching Strategies GOLD® subscales (Social-Emotional, Cognitive, Literacy, and Mathematics) and kindergarten readiness (Not Ready, In Process, and Proficient) moderated by gender (female vs. male)?

RQ5. To what extent are the relationships between Teaching Strategies GOLD® subscales (Social-Emotional, Cognitive, Literacy, and Mathematics) and kindergarten readiness (Not Ready, In Process, and Proficient) moderated by ethnicity (Hispanic vs. Non-Hispanic)?

Research Design

This study was quantitative and non-experimental in nature. The study used existing data (secondary analysis) from two cohorts of preschool children enrolled in a blended Head Start/state-funded voluntary prekindergarten program (VPK). The independent variables (VPK

Assessments, Teaching Strategies GOLD®, gender, and ethnicity) were ascertained during the spring of the blended Head Start/VPK program year. The dependent variable, Work Sampling System™ (WSS), was administered within the first 30 student contact days of the children’s kindergarten year; approximately four months after the independent variable data were collected for both cohorts.

Data Sources

Florida VPK Assessment Measures. The Florida VPK Assessment measures (FLDOE, 2011) were created at the Florida Center for Reading Research (FCRR) under the supervision of Dr. Christopher Lonigan. The Florida VPK Assessment measures were designed to assess four distinct skill domains: print knowledge, phonological awareness, mathematics, and oral language/vocabulary skills. The VPK Assessment Measures were parallel in form, designed for children to be assessed individually by the child’s VPK teacher within three finitely assigned assessment periods (AP): (a) AP 1 – month of September; (b) AP 2 - month of January; and (c) AP 3 – month of May. Although the assessment time for each child may vary from a minimum of 15 minutes, FCRR suggested that teachers administer the four skill domains to each child in the following order: (1) Print Knowledge; (2) Phonological Awareness; (3) Mathematics; and (4) Oral Language/Vocabulary Skills (FLDOE, 2011). FCRR determined this order of skills based on literature that stated that Print Knowledge skills are often easier for young children to master. According to Lonigan (FLDOE, 2011), each of the three parallel forms of the VPK Assessment Measures is designed to assess the range of abilities within each of the four skill domains that are developmentally appropriate with that of a skill likely to be exhibited by children during their 4-year-old preschool year.

The authors of the Florida VPK Assessment measures stated that the goal of the measures was to provide VPK teachers and other professionals with a means of identifying children who are not on a trajectory of success to exhibit kindergarten readiness in terms of their developing reading-related and math skills during the children's VPK experience. The VPK Assessment was created through an iterative process of item development, field testing of items, and item refinement that was presented in a field trial of the assessment measure by VPK teachers. Initial item development involved creating sets of items that mapped onto the domains of early literacy and early math skills that are included in *Florida's Early Learning and Developmental Standards for Four-Year-Olds* (FLDOE, 2011b). Within each skill domain, the authors created items using different formats to identify a variety of means of assessing children's skills, as well as a manner to provide a range of questioning/response formats (i.e., question stems, response formats).

In order to determine score reliability within the VPK Assessment, a combination of Item Response Theory (IRT) and Classical Test Theory (CTT) was utilized in the domains of print knowledge, phonemic awareness, mathematical skills, and oral language. As the purpose of the VPK Assessment was to identify young children who may be at risk of not meeting identified criteria for kindergarten readiness, items were selected by the authors to increase precision around scores likely to represent a higher region of risk (FLDOE, 2011). Therefore, the authors expected that the standard errors would be more precise (lower) for theta values (i.e., ability) from average to below average than for theta values in the above average range (FLDOE). The IRT analysis revealed that each of the three versions encapsulating the four domains provided a high degree of precision of measurement in the region of the ability distribution most relevant for identifying children who have weak early language, literacy, or math skills. For all measures,

but particularly for the print knowledge, oral language, and math measures, precise measurement was obtained over a wide range of abilities that spanned from around average to well below average levels.

Classical Test Theory (CTT) analyses of the VPK Assessment related to the internal-consistency reliability (coefficient alpha) were examined for each version of individual measures within the three assessment periods (AP1, AP2 and AP3). Coefficient alphas were not only reviewed for each measure but for each sample used in the development process (i.e., development, validity, teacher field trial, field trial retest, and all samples combined). All samples of each measure of the four skill areas showed moderate to high levels of internal consistency within several independent samples. Table 1 provides the internal consistency ranges for each of the four measures and samples for Assessment Period 3 only as this assessment window is examined within the current study.

Table 1
Internal Consistency Reliability Ranges for VPK Assessment Measures (AP3) for Samples Used in Different Phases of VPK Measure Development

Measure	Internal Consistency Reliability Ranges
Print Knowledge	.84 - .90
Phonological Awareness	.82 - .88
Mathematics	.79 - .94
Oral Language/Vocabulary	.70 - .86

Note. Approximate *N* in Sample: *Development* *N*=512; *Validity* *N*=302; *Teacher Field Trial* *N*=1,075 to 1,227; *Field Trial Retest* *N*=124 to 146; *All Samples Combined Development* *N*=2,013 to 2,189 (FLDOE, 2011)

Analyses of alternate-forms reliability demonstrated that the three versions of each measure were assessing the same underlying ability and provided strong support for using the VPK Assessment's three parallel forms (i.e., AP1, AP2 and AP 3) as a means to monitor children's development of skills in the four skill domains. Finally, analyses of test-retest reliability indicated that each measure had moderate to high levels of cross-time stability at a level expected given the measures' internal-consistency reliabilities and the narrow length of time (two to three month intervals) between assessments in the test-retest analyses (FLDOE, 2011).

Validity coefficients between scores on the VPK Assessment and the scores on the two components of the version of the Florida Kindergarten Readiness Screener (FLKRS) that were utilized at the time of this validation process, Early Childhood Observation Screener (ECHOS) and Florida Assessment for Instruction in Reading for Kindergarten (FAIR-K), were moderate ($r = .20$ to $.32$ and $r = .31$ to $.55$, respectively). However, with only a few exceptions, these correlations were similar to the correlations between scores on the ECHOS and scores on the FAIR-K measures that were administered concurrently, rather than 12-, 7- or 5-months before the ECHOS as was the case with the VPK Assessment. Despite this limited variability and the absence of a clear pattern of item covariance that would indicate that the items on the ECHOS assess different constructs, there was some evidence of the expected pattern of convergent and discriminant relations (FLDOE, 2011). All but one of the convergent correlations for the VPK Assessment was significantly higher than the correlations between the VPK Assessment and the socio-emotional scale constructed from the ECHOS. The overall pattern of results for the VPK Assessment was similar to the pattern of results for the FAIR-K measures, which suggests that the VPK Assessments were at least as valid with respect to the ECHOS and the FAIR-K

measures (FLDOE, 2011). Table 2 represents the Florida VPK Assessment subscale methodology, variable, and subscale types.

Teaching Strategies GOLD® Assessment. *Teaching Strategies GOLD®* online assessment system (Heroman, Burts, Berke, & Bickart, 2010) is an authentic, observation-based performance assessment system of young children. Authentic assessment systems are ongoing with teachers/caregivers gathering daily anecdotal data within daily routines and settings rather than an additional data collection cycle or source. *Teaching Strategies GOLD®* has 36 objectives organized within the areas of approaches to learning, language, cognition, literacy, mathematics, and physical and social-emotional development. Teachers gather student anecdotal data within daily activities as the teachers observe and interact with children and families (Heroman, Burts, Berke, & Bickart). In an effort to determine the concurrent validity of *Teaching Strategies GOLD®*, an independent not-for-profit organization, American Institute for Research (AIR), examined the associations between the *Teaching Strategies GOLD®* scale scores and (1) teacher ratings of children’s social functioning and learning behaviors and (2) child performance on direct assessments of academic skills. AIR selected six widely used external direct assessment measures exhibiting strong psychometric properties of reliability and validity, and which align closely to the *Teaching Strategies GOLD®* domains. The direct assessment instruments selected for this study were: (a) Pre-Language Assessment Scales (Pre-LAS) (Duncan & De Avila, 1985); (b) Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4) (Dunn & Dunn, 2007); (c) Woodcock-Johnson Tests of Achievement (WJ-J III) (Woodcock, McGrew, & Mather, 2007); (d) Preschool Self-Regulation Assessment (PSRA) – Pencil Tapping Task only (Smith-Donald, Raver, Hayes, & Richardson, 2007); (e) Head-Toe-Knees-Shoulders Task (HTKS) (Ponitz, McClelland, Matthews, & Morrison, 2009); (f) Preschool and

Table 2

Florida VPK Assessment General Outcomes Subscale Variables for Kindergarten Readiness

Measure	Description	Method	Subscale Type
<i>Print Knowledge</i>	Measures child's ability to recognize letters and words, as well as the sounds they make	12 assessment items and 2 practice items	Continuous: 0 - 12
<i>Phonological Awareness</i>	Measures child's ability to: blend a word if it is broken into smaller sounds or syllables; blend a compound word; recognize the rest of the word when part of the stimulus is taken away	14 items and 2 practice items	Continuous: 0 - 14
<i>Mathematics</i>	Measures child's early numeracy skills across three areas: counting; numerical relations; and arithmetic reasoning	13 items and 0 practice items	Continuous: 0 – 18*
<i>Oral Language/ Vocabulary</i>	Measures child's expressive/receptive language and knowledge of adjectives, verbs, verb tenses, prepositions and nouns	23 items and 0 practice items	Continuous: 0 - 23

Note. Variable type is the same within each subtest. Each item uses a binary response of 'Correct' or 'Incorrect'

* Item 1 is a 5-row grid with each correct row worth 1 point; this makes the range for Mathematics 0 to 18

(g) Preschool Learning Behaviors Scale (PLBS) (McDermott, Leigh, & Perry, 2002).

Over a period of one month, trained AIR data collectors individually administered the six direct assessment measures, while classroom teachers who were current *Teaching Strategies GOLD*® users collected assessment information, as well as the measures for social functioning and learning skills. Within this concurrent validation study, a stratified random sample of Head Start, private and public preschool pre-k and other types of early childhood providers were utilized. The sample consisted of 299 children within the northeastern United States nested within 51 different classrooms nested within 16 centers. The sample was equally distributed by sex and a majority of the children (59%) lived within a home where English was the primary language spoken. Nearly 25% of the sample was Dual Language Learners who could be assessed and respond in English and about 25% of the children were from low-income families. The sample had an ample distribution of ethnicities/races (Hispanic: 45%; African American: 26%; White: 29%).

This study began by examining Pearson correlation coefficients to assess the degree of association between the external measures and the *Teaching Strategies GOLD*® scale scores. However simple Pearson correlations could not account for the fact that children were nested within classrooms or that the classroom teachers completed both the *Teaching Strategies GOLD*® and PKBS and PLBS assessments. In order to account for these situations, a two-level hierarchical linear model (HLM) was used to evaluate each correlation coefficient between each external measure and *Teaching Strategies GOLD*® scale score (i.e., Language, Literacy and Mathematics). The results from these models were used to estimate the variance in the external measures accounted for by the individual *Teaching Strategies GOLD* scale score while accounting for teacher ratings and clustering effects (Teaching Strategies, 2013). Findings from

the study found the correlations of Teaching Strategies GOLD® and scale scores of the external measures were moderate within aligned areas. It was determined that the scale scores from each external measure were associated with the Teaching Strategies GOLD® scales scores measuring the most closely related construct and resulted in the strongest correlations. Table 3 illustrates the ranges of correlations with the specific Teaching Strategies GOLD® scale scores.

Table 3
Pearson Correlation Ranges between External Measures and Teaching Strategies GOLD® Scale Scores

External Measure	Teaching Strategies GOLD® Domains with Moderate Correlations	Pearson Correlation Range
Peabody Picture Vocabulary Test (PPVT™ - 4)	Language, Literacy, Mathematics, Cognition	$r = .44$ to $.48$
Head – Toe – Knees – Shoulders Task (HTKS)	Language, Literacy, Mathematics, Cognition	$r = .36$ to $.39$
Pre-Language Assessment Scales (Pre-LAS)	Social-Emotional, Physical, Language, Literacy, Mathematics, Cognition	$r = .31$ to $.41$
Pencil Tapping Task of the Preschool Self-Regulation Assessment (PSRA)	Social-Emotional, Physical, Language, Literacy, Mathematics, Cognition	$r = .37$ to $.48$
Woodcock-Johnson® III Tests of Achievement (W-J III) Letter-Word Identification Word Attack Understanding Directions	Literacy	$r = .37$ to $.45$
Woodcock-Johnson® III Tests of Achievement (W-J III) Quantitative Concepts	Social-Emotional, Physical, Language, Literacy, Mathematics, Cognition	$r = .40$ to $.52$

Note. Approximate N in Sample: 299 (Teaching Strategies, 2013)

Within the hierarchical linear models, for almost every external direct assessment measure, the *Teaching Strategies GOLD®* Literacy scale score showed the strongest association

among the scale scores. The literacy scale score yielded a moderately strong association (10.4% to 39.7% variance accounted for) with the strength of association very similar to or higher than that shown with the Pearson correlation coefficient. These results made sense as a majority of the direct assessment measures focused on either literacy-related constructs or a significant amount of literacy related components. Some of the associations became notably higher within the hierarchical linear models than by examining the simple Pearson correlation coefficients (Teaching Strategies, 2013). The overall assessment measures the knowledge, skills, and behaviors that are predictive for school success (Heroman, Burts, Berke, & Bickart, 2010). Twenty-three objectives have been organized into six areas of development and learning: (1) Social-Emotional; (2) Physical; (3) Language; (4) Cognitive; (5) Literacy; and (6) Mathematics. Based on the literature related to predicting kindergarten readiness as well as the subscales comprising the Florida VPK Assessments, the current study focused on only four subscales, not examining relationships with the Physical subscale. Within each subscale, ordinal scale scores (ranging from 0 to 9) are calculated into one of three categorical scores: Below Expectations, Meeting Expectations, and Exceeding Expectations. Table 4 displays the four curriculum-embedded subscales of Teaching Strategies GOLD® with subscale descriptions, method of gathering assessment data as well as variable type.

Work Sampling System® P4, 5th Edition. The Work Sampling System (WSS) (Meisels, Jablon, Dichtelmiller, Marsden, & Dorfman, 2001) is an observational authentic assessment for children from preschool (age 3) through Grade 6. The publisher states that the assessment has been used with more than 850,000 children, mostly in prekindergarten and kindergarten enrolled in nearly 10 states, New York City and most recently, Florida (Meisels, 2011). WSS is a criterion-based, curriculum-embedded performance.

Table 4
Teaching Strategies GOLD® Curriculum-embedded Subscale Variables for Kindergarten Readiness

Measure	Description	Method
<i>Social-Emotional</i>	Regulate own emotions Establish positive relationships Participates in group situations	9 items
<i>Cognitive</i>	Approaches to learning Remembers and connects experiences Classification skills Uses symbols for items not present	10 items
<i>Literacy</i>	Phonological awareness Alphabet knowledge Concepts of print Responds to books and other texts Emergent writing skills	12 items
<i>Mathematics</i>	Number concepts and operations Spatial relationships and shapes Compares and measures	7 items

Note. Teaching Strategies GOLD® variable types for each item are ordinal from 0 – 9.
 Subscale types for each measure are continuous summary scores

assessment that was created to document and determine the state of what children are learning and have begun to master by providing information on individual students' academic and social achievements (Meisels). The measure allows teachers to collect a wide-span of information from multiple sources and focus the observations to evaluate what children know and can do during the specific assessment window. Customized versions of the WSS have been created by the publisher, NCS Pearson for State Education Agencies (SEAs), Local Education Agencies (LEAs), and Head Start.

This study focused on the WSS P4 5th edition utilized within the state of Florida as a state-mandated kindergarten readiness screener for all first time, non-retained kindergarten students. This WSS edition is comprised of 47 performance indicators within five domains: Personal and Social Development; Language and Literacy; Mathematical Thinking; Scientific Thinking; and Physical Development, Health, and Safety. Performance indicators describe the skills, behaviors, and accomplishments representing end-of-year expectations for four-year-old children. FLKRS performance indicators align closely with the *Florida Early Learning and Developmental Standards for Four-Year-Olds*. Each performance indicator was written as one sentence in order to help teachers understand and document students' performance (FLDOE, 2016). These guidelines foster consistency of interpretation and evaluation among different teachers, children, and schools (Meisels, 2011). Teachers rate a child's performance on each categorical performance indicator as either *Not Yet*, *In Process*, or *Proficient*.

The Florida Department of Education provided operational definitions of *Not Yet*, *In Process*, or *Proficient* on the WSS within the 2015-2016 Florida Kindergarten Readiness Screener Administration Manual (FLDOE, 2015) with the operational definition of *Not Yet* indicating that the performance indicator represents a skill, an area of knowledge, or a specific set of behaviors or accomplishments that the child has not acquired. The operational definition of *In Process* signifies that the skills, knowledge, behaviors, or accomplishments described by the performance indicator are intermittent or emergent, and are not demonstrated reliably or consistently. The operational definition of *Proficient* indicates the child's skill, knowledge, or behavior matches the end-of-year expectations described in the rationales of the WSS developmental guidelines.

A study focused on the reliability and validity of the WSS was conducted with 100 kindergarten-aged children nested within 10 classrooms in three Michigan school districts. The study authors created a classical psychometric design whereby the children were enrolled within classrooms implementing the WSS and were also given individually-administered norm-referenced assessments in the fall and spring of the students' kindergarten year (Meisels, Liaw, Dorfman, & Fails, 1995). Throughout the school year, the teachers completed the WSS checklist, which included 69 items within five domains three times a year (fall, winter, and spring). The domains were: (1) art and fine motor; (2) movement and gross motor; (3) concept and number; (4) language and literacy; and (5) personal/social development. The teacher selected the following rating of the student's performance on each item (1) not yet; (2) sometimes; (3) often. At the end of the school year, the teacher selected a composite score of the child's performance across each of the five domains as (1) not yet accomplished; (2) accomplished; or (3) highly accomplished. The reliability of the WSS was examined by Cronbach alphas and correlations (Meisels, Liaw, Dorfman, & Fails). A subscale score was created for each domain by summing the individual items of the domain. A total score for the developmental checklist in the fall, winter, and spring was then computed by adding each of the five subscales. The correlations indicated a moderate to high level of reliability of measurement across the school year, with the largest between the fall and winter checklists and winter and spring checklists, .89 for both time periods. The degree of the internal consistency among items for the five domains at all three checkpoints were determined by calculating Cronbach's alphas. High levels of internal consistency of the checklists were noted with alphas ranging from .87 to .94 (Meisels, Liaw, Dorfman & Fails).

Two individually-administered norm referenced assessments were given to the students in the fall and spring. Six subtests of the Kindergarten Achievement Battery of the Woodcock-Johnson Psychoeducational Battery-Revised (WJ-R; Woodcock-Johnson, 1989) were administered. A total score for the WJ-R was derived by summing the standard scores of the six subtests, letter word identification, applied problems, dictation, science, social studies and humanities.

The second assessment was the Motor Scale of the McCarthy Scales of Children's Abilities (MSCA; McCarthy, 1972). These assessments were administered by five trained examiners who were blind to the study's purpose. The order of the assessment in the fall and spring were counter balanced to avoid order effects. Index scores were used to determine the student's motoric development (Meisels, Liaw, Dorfman, & Fails, 1995).

The researchers utilized zero-order correlations between the WSS and the two assessments (WJ-R and MSCA) to determine concurrent validity. Moderate to high correlations were found between the WSS and the WJ-R ($r = .75$ for the fall and $r = .66$ for the spring). However, the correlations between the WSS and MSCA were low ($r = .39$ for the fall and $r = .28$ for the spring). The predictive validity of the WSS was also examined with the means of correlations, regression and computation of sensitivity and specificity relating to the fall and winter WSS to the spring administration of the WJ-R and MSCA (Meisels, Liaw, Dorfman, & Fails, 1995). As with the concurrent validity, high correlations were found between the fall and winter WSS and spring WJ-R ($r = .66$ and $.76$). However, the correlations were moderate to low between the WSS and spring MSCA ($r = .43$ and $.34$) (Meisels, Liaw, Dorfman, & Fails).

Measure Procedures

VPK Assessment measures addressing print knowledge, phonological awareness skills, mathematics skills and oral language skills and the Teaching Strategies GOLD® measures addressing approaches to learning, general cognition (including mathematics), language and literacy, physical health and social-emotional skills were administered to each student across the preschool academic year. Training through professional development sessions with a state-certified trainer was provided for Head Start teachers and coaches to administer the VPK Assessment to all four-year-old Head Start children (FLDOE, 2011). VPK Assessments were administered within the month of September (AP1) and April/May (AP3) as directed by the Florida Department of Education's Office of Early Learning Department. As the current investigation was intended to address the predictive validity of the VPK Assessment for school readiness as determined by performance on the FLKRS Work Sampling System, measures from AP 3 of the VPK Assessment were the primary focus of this investigation.

With regard to the Teaching Strategies GOLD® (Heroman, Burts, Berke, & Bickart, 2010), Head Start teachers were provided with a three-day, face-to-face professional development session by a company representative. Additional professional development opportunities were provided to Head Start teachers by participating in and passing four online professional development modules. Teaching Strategies GOLD data were collected in the fall (October), winter (January) and spring (April/May). Head Start coaches and administrators observed teachers' Teaching Strategies GOLD® data collection at the winter and spring administration checkpoints to determine inter-observer agreement. As the current investigation was intended to address the predictive nature of the Teaching Strategies GOLD® on school

readiness as determined by performance on the FLKRS Work Sampling System, checkpoint data from the spring administration cycle were the primary focus of this investigation.

During the first 30 student contact days of kindergarten, district kindergarten teachers administered the FLKRS Work Sampling System (i.e., Approaches to Learning, Print Knowledge, Language and Literacy, Early Mathematic and Social-Emotional). Categorical data (i.e., Proficient, In Process and Not Yet) for each performance indicator of the Work Sampling System were entered into an electronic repository hosted by the Florida Center for Reading Research (FCRR) at Florida State University in Tallahassee, Florida. Once the WSS data has been entered into the data repository, numeric points are given to each performance indicator of the WSS based on the categorical data selected. One point was provided for each performance indicator with *Not Yet* selected, two points were provided for each performance indicated with *In Process* selected and three points were provided for each performance indicator with *Proficient* selected. Pearson testing company calculated and aggregate numerical score and then reformulated the composite scores back into categorical scores of *Not Yet*, *In Process*, and *Proficient* based on the number of points and number of items the child was rated on the WSS. Appendix B contains WSS score range tables provided by the Florida Department of Education (2014).

Participant Selection

The participants in this study were a convenience sample given that they were readily available by the school district with a large sample of preschool aged children enrolled within a blended Head Start/VPK program. The sample consisted of two cohorts of four-year-old children enrolled in a blended Head Start/VPK program in 2014-2015 and 2015-2016 within a large county-based school district in Florida. The school district is located within a 745 square

miles of land area that contains a mix of both rural and suburban communities on the west coast of Florida, containing approximately 2.5% of the state's citizens ($N = 475,502$) (U.S. Census, 2015). Five percent of the county's population is children under five years of age ($N = 23,775$) with nearly 3,305 of those children living below the poverty line. Ninety percent of the ethnic demographic of the county is White, Non-Hispanic with the following disaggregated estimates by race: (1) African-American, 5%; (2) Asian, 2%; (3) Caucasian, 78%; (4) Hispanic, 13%; and (5) Multi-racial 2% (U.S. Census).

Data consisted of scores from the Florida Voluntary Prekindergarten (VPK) Assessments (FLDOE, 2011) and the Teaching Strategies GOLD®. The researcher asked permission from the University's Institutional Review Board and the school district's Accountability and Research Department to gain access to the VPK Assessments and Teaching Strategies GOLD® scores from the sample of preschool-aged children enrolled within the federal income-based Head Start program and the universal state voluntary prekindergarten program (Florida VPK). Children from birth to five years from families below the federal poverty guidelines were eligible to participate in Head Start, while the state funded universal prekindergarten was available for any child who was four years of age on or before September 1 of the school year. These early childhood programs were comprised of nearly 700 children who were nested within 39 classrooms nested within 22 elementary schools. A majority of these schools received federal Title I funding to offset academic/social challenges attributed to enrolling large percentages of children from low-socio-economic households. Tables 5a and 5b graphically display the early childhood blended Head Start/VPK program with children nested within 22 schools and nested within 39 classrooms with the sample of children who participated in the VPK Assessment and Teaching Strategies GOLD for the 2014-15 and 2015-2016 cohorts.

Table 5a

Distribution of ECE Children Nested Within Classrooms Within Schools with Completed VPK Assessments

School	Teacher <i>N</i>	Cohort 1: 2014-15 Children with Completed VPK Assessment	Cohort 1: 2015-16 Children with Completed VPK Assessment
A	2	32	31
B	1	14	16
C	1	15	15
D	2	28	27
E	2	32	27
F	2	33	28
G	2	30	25
H	1	13	19
I	1	14	13
J	2	30	32
K	2	31	28
L	1	13	17
M	2	29	30
N	2	29	24
O	2	34	33
P*	1	15	15
Q	2	32	27
R	4	54	50
S	1	17	15
T*	2	20	26
U	2	30	34
V	2	30	24

Note. * - Non-Title I funded schools; all other schools receive federal Title I funds

Data from the 1,516 children enrolled within the two cohorts (2014-2015 cohort $n = 747$ and 2015-2016 cohort $n = 769$) were reviewed to determine how many children met inclusion and exclusion criteria for the study. Children who did not complete the early childhood program year or did not have valid VPK Assessments and Teaching Strategies GOLD® assessment data were excluded from this sample.

Table 5b

Distribution of ECE Children Nested Within Classrooms Within Schools with Completed Teaching Strategies GOLD® Assessments

School	Teacher <i>N</i>	Cohort 1: 2014-15 Children with Completed GOLD Assessment	Cohort 1: 2015-16 Children with Completed GOLD Assessment
A	2	29	29
B	1	12	16
C	1	13	15
D	2	25	23
E	2	30	25
F	2	29	25
G	2	29	23
H	1	10	17
I	1	14	10
J	2	30	32
K	2	28	24
L	1	11	16
M	2	26	25
N	2	27	23
O	2	29	33
P*	1	15	14
Q	2	27	23
R	4	45	44
S	1	15	15
T*	2	15	25
U	2	27	35
V	2	26	25

Note. * - Non-Title I funded schools; all other schools receive federal Title I funds

Children who did not have valid FLKRS kindergarten readiness data in 2015-2016 and 2016-2017, respectively, were also excluded from the final sample.

Race was defined by four categories: (1) White; (2) African American; (3) Native American, Native Alaskan; and (4) Asian Pacific Islander (U.S. Census, 1999). Ethnicity was defined in two categories: (1) Hispanic and (2) Non-Hispanic (U.S. Census). Years of participation was operationally defined by the number of years that the child had been enrolled in Head Start. In this school district, less than 15% of the children selected to participate in the

Head Start program are three years of age. Therefore, a participation year of one was operationally defined as children who were enrolled in Head Start at four years of age and a participation year of 2 included children who were enrolled in Head Start at three years of age. Only children with a participation year of one were included in the current study. In order to determine the primary language spoken at home, parents were required to complete a district-created Home Language Survey. Language spoken at home was defined as English or Spanish.

Data Analysis

Prior to using the existing demographic and assessment data, district and university level permission was obtained by submitting the proposal to the school district's Office of Accountability and Research and the University's Institutional Review Board (IRB). Once approval was obtained, the researcher, who also serves as the Quality Assurance Manager of the Head Start program, retrieved the data. The information was gathered from files previously submitted to the state and analyzed using SPSS version 21 software.

Variables. The student level demographic variables included student gender and student ethnicity. The VPK Assessment variables included Print Knowledge, Phonological Awareness, Mathematical Concepts and Oral Language and Vocabulary. The Teaching Strategies GOLD® assessment variables included Social-Emotional, Cognition (focusing on approaches to learning), Literacy, and Mathematics. The VPK Assessments subscale variables were continuous summary scores calculated from binary correct (1) and incorrect (0) scores within each subtest, while the Teaching Strategies GOLD® assessment variables were continuous summary scores calculated by scores from each dimension (ordinal scale scores of 0 to 9).

When examining the relationships between the VPK Assessments and Teaching Strategies GOLD® and the kindergarten readiness of the child, the VPK Assessments and Teaching Strategies GOLD® served as the predictor variables, respectively, while the criterion measure was the kindergarten readiness score. Table 6 represents the variables and a description of each.

Preliminary Analysis. Descriptive analyses were used to describe the data set, as well as determine the distribution, central tendency and variation of the variables. Although logistic regression does not have the same assumptions of normality as other statistical procedures, preliminary analyses were conducted in an effort to screen for missing data across variables and determine any relationships missing data may have with other variables.

Statistical Analysis. As the dependent variable of kindergarten readiness was categorical and ordinal in nature, multiple ordinal logistic regression analyses were conducted to determine the predictive ability of the VPK Assessments and Teaching Strategies GOLD® on kindergarten readiness. Categorical values on the Work Sampling System were defined as *Not Yet* (category 0), *In Process* (category 1), and *Proficient* (category 2). The goal of these multiple ordinal logistic regression models was to examine the relation between kindergarten readiness and several assessment measures. Results were reported as the regression coefficient (β), standard error of the regression coefficient ($SE \beta$), Odds Ratio (e^β), Wald test (χ^2), degrees of freedom (df), significance, and confidence intervals. The models were also examined for Goodness of Fit using Maximum Likelihood Ratios and R^2 .

Table 6

Description of Independent Variables in the Multiple Logistic Regression Models

Variable	Description
<i>Student Demographic Control Variables</i>	
Gender	0 - Females 1 - Males
Ethnicity	0 – Hispanic 1 - Non-Hispanic
<i>Predictor Variables</i>	
<i>VPK Assessments</i>	
Print Knowledge	Summary Score Range: 0 - 12
Phonological Awareness	Summary Score Range: 0 - 14
Mathematics	Summary Score Range: 0 - 18
Oral Language/Vocabulary	Summary Score Range: 0 – 23
<i>Teaching Strategies GOLD®</i>	
Social - Emotional	Summary Score Range: 0 - 81
Cognitive	Summary Score Range: 0 - 90
Literacy	Summary Score Range: 0 - 108
Mathematics	Summary Score Range: 0 - 63
<i>Outcome Variable</i>	
<i>Work Sampling System™</i>	
Personal and Social Development	Summary Score Range: 12 – 36
Language and Literacy	Summary Score Range: 10 – 30

Table 6 Continued

Mathematical Thinking	Summary Score Range: 12 – 36
Scientific Thinking	Summary Score Range: 4 – 12
Physical Development, Health, and Safety	Summary Score Range: 7 – 21

As this secondary analysis consisted of a sample of children nested in classrooms, *Mplus* (Muthén & Muthén, 1998-2017) was utilized to adjust for the standard errors presented within the models. The Huber-White sandwich estimator was utilized in each model to estimate the variance of the Maximum Likelihood Ratios if the underlying models were incorrect (Freedman, 2005). This sandwich estimator was added into the *Mplus* model syntax with the command `Analysis: type = complex`. An example of the complete *Mplus* syntax for a cohort 1 model examining the relationship between the VPK Assessment subscale Print Awareness and gender can be found in Appendix C. The first set of analyses focused on each assessment subscale, one at a time, as a predictor of kindergarten readiness and evaluated the robustness of the relationships. I looked at one main effect subscale at a time. I then evaluated the potential interaction of each subscale with gender first and then with ethnicity. For a relationship to be considered robust, the relationship for each subscale had to be statistically significant ($p \leq .01$) for both cohorts.

The second set of analyses simultaneously included all of the subscales within an assessment (e.g., VPK Assessment) and were used to evaluate the relationship between a subscale and the outcome, statistically controlling for the other subscales within the assessment (e.g., VPK Assessment). As with the previous models, for a relationship to be considered robust, the relationship for each subscale had to be statistically significant ($p \leq .01$) for both cohorts.

A series of models were run starting with Model 1 that consisted of the control variables (gender and ethnicity), Model 2 consisting of control variables and main effects (VPK Assessment subscales), and then Models 3a – 3d consisting of control variables, main effects for VPK subscales, and interactions between the control variable of gender and each individual subscale. The same modeling series, Models 4a – 4d, was used to examine the relationships between the control variables, main effects of VPK Assessment subscales, and the interaction between ethnicity and individual VPK Assessment subscales. A series of models, following the same strategy described above, was used to examine the Teaching Strategies GOLD® subscales. Tables containing the specific model analyses are located in Appendix D.

In total, 98 models were run to determine the predictive and differential validity of the VPK Assessment subscales and Teaching Strategies GOLD® subscales on the FLKRS Work Sampling System™ (WSS). As the number of models increases, so does the Type I error rate, the confidence level for the statistical models was increase from 95% to 99%.

Chapter Summary

In this chapter I described the purpose of the study, the research questions, and the sample. The purpose of this study was to investigate the relations of the standards-based assessment measure (VPK Assessment) and the curriculum-embedded assessment measure (Teaching Strategies GOLD®) and assess their predictive validity of school readiness at kindergarten entry with the Florida Kindergarten Readiness Screener (FLKRS) based on the Work Sampling System (WSS). The main statistical approach to assess the predictive validity was ordinal logistic regression as the outcomes score was a categorical variable.

CHAPTER FOUR

RESULTS

The purpose of this quantitative correlational study was to understand the predictive validity of the Florida VPK Assessments and Teaching Strategies GOLD within classrooms in schools in a blended Head Start/VPK program in Florida. The study focused on determining if predictive relationships existed between the independent variables (i.e., VPK Assessments and Teaching Strategies GOLD®) and the dependent variable of kindergarten readiness. Additionally, the study examined whether selected child characteristics of gender and ethnicity moderated the predictive relationships between the independent variables of the subscales from the VPK Assessments and Teaching Strategies GOLD® assessments and the dependent variable of kindergarten readiness.

In this chapter, I present the results of analyses described in Chapter Three. Data collection and descriptive statistical analyses for each cohort are illustrated and discussed. I will complete the chapter with the results of the main statistical analyses for each cohort used to answer the research questions.

Data Collection

As this study was a secondary analysis, data consisted of scores from the Florida Voluntary Prekindergarten (VPK) Assessments (FLDOE, 2011) as the standardized measure and Teaching Strategies GOLD® as the curriculum-embedded, authentic measure. The researcher received permission from the University's Institutional Review Board and the school district's Accountability and Research Department to gain access to the VPK Assessments and Teaching

Strategies GOLD® scores from the sample of preschool-aged children enrolled within the federal income-based Head Start program and the universal state voluntary prekindergarten program (Florida VPK). Children from birth to five years from families below the federal poverty guidelines are eligible to participate in Head Start, while the state funded universal prekindergarten is available for any child who is four years of age on or before September 1 of the school year. These early childhood programs were comprised of children who were nested within 39 classrooms nested within 22 elementary schools. A majority of these schools received federal Title I funding to offset academic/social challenges attributed to enrolling large percentages of children from low-socio-economic households. Cohort 1 children were enrolled within the blended Head Start/VPK classrooms during the 2014-2015 academic year, while Cohort 2 children were enrolled during the 2015-2016 academic year. Cohort 1 had a sample of 604 children with nearly 48% of the sample being female and nearly 38% of the sample indicated their ethnicity was Hispanic. Cohort 2 had a sample of 565 children with nearly 52% females and 33% percent Hispanic. The overall combined sample of both cohorts was 1,169 children with nearly 50% females and 35% with an identified ethnicity of Hispanic. Table 7 represents the demographic data with relation to gender and ethnicity of Cohorts 1 and 2.

Descriptive Statistics

Descriptive analyses were used to describe the data set, as well as determine the distribution, central tendency, and variation of the variables. Tables 8 through 15 present descriptive statistics for the entire cohort by the independent variables: VPK Assessment subscales, Teaching Strategies GOLD® subscales, control variables of gender and ethnicity (Hispanic and Non-Hispanic) and the categorical dependent variable FLKRS Work Sampling

Table 7

Preliminary Demographic Information of the Sample Cohorts

	Year	Sample	Gender (Female of Cohort)	Ethnicity (Hispanic of Cohort)
	Spring of Head Start	Kindergarten	<i>N</i>	<i>N</i> (%)
Cohort 1	2015	2015 - 2016	604	288 (47.7)
Cohort 2	2016	2016 - 2017	565	293 (51.9)
Total	--	--	1,169	581 (49.7)

System™ (WSS). Both Cohort 1 (Tables 8 through 11) and Cohort 2 (Tables 12 through 15) exhibited the highest mean scores within VPK Assessment on the Oral Language/Vocabulary subscale ($M = 18.23$, $SD = 3.51$ and $M = 18.70$, $SD = 6.54$, respectively). The highest mean score for both cohorts for Teaching Strategies GOLD® was on the Literacy subscale ($M = 61.67$, $SD = 14.19$ and $M = 62.71$ and $SD = 13.76$, respectively). The subscales had skewness values that were within the acceptable range of +/- 2.0. With the exception of the Cognitive subscale of Teaching Strategies GOLD®, all subtests were negatively skewed for both cohorts. The VPK Assessment subscales and all Teaching Strategies GOLD® with the exception of the Social-Emotional scale had distribution of kurtosis that were platykurtic with shorter and thinner tails and broader central peaks. However, the Social-Emotional subscale had a distribution of kurtosis larger than 3.0 indicating a leptokurtic distribution with longer and fatter tails with a higher and sharper central peak (Westfall, 2014).

Table 8

Descriptive Statistics: Full Sample for Cohort 1

Variable	<i>N</i>	<i>M</i>	<i>SD</i>	Range	Skewness	Kurtosis
VPK Assessment – AP3						
Print Knowledge	575	9.23	3.16	0-12	-0.95	-0.33
Phonological Awareness	575	9.34	3.50	0-14	-0.45	-0.78
Mathematics	575	12.06	4.22	0-18	-0.61	-0.43
Oral Language/Vocabulary	575	18.23	3.51	0-23	-1.33	2.52
Teaching Strategies GOLD – Spring						
Social – Emotional	533	57.31	8.14	33-81	-0.20	-0.43
Cognitive	533	59.90	8.81	31-88	0.10	0.72
Literacy	530	61.67	14.19	20-103	-0.10	-0.05
Mathematics	531	32.81	6.30	10-49	-0.43	0.37

Note. The potential range each subscale of the VPK Assessment is consistent with the actual range listed within the table

Table 9

Descriptive Statistics: Gender Sample for Cohort 1

Variable	Gender									
	Female (0)					Male (1)				
	<i>N</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis	<i>N</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
VPK Assessment – AP3										
Print Knowledge	274	9.56	2.89	-1.07	0.05	301	8.92	3.37	-0.81	-0.66
Phonological Awareness	274	9.51	3.40	-0.45	-0.81	301	9.18	3.59	-0.44	-0.77
Mathematics	274	12.42	3.84	-0.60	2.08	301	11.74	4.52	-0.56	-0.61
Oral Language/Vocabulary	274	18.42	3.37	-1.31	-0.36	301	18.07	3.63	-1.33	2.78
Teaching Strategies GOLD - Spring										
Social - Emotional	265	58.65	7.85	-0.33	-0.18	268	55.99	8.23	-0.05	0.48
Cognitive	265	60.78	8.32	-0.04	0.59	268	59.03	9.20	0.25	0.89
Literacy	262	63.47	13.07	-0.11	0.14	268	59.02	15.01	-0.01	-0.20
Mathematics	263	33.46	5.66	-0.30	-0.14	268	32.17	6.82	-0.42	0.38

Table 10

Descriptive Statistics: Ethnicity Sample for Cohort 1

Variable	Ethnicity									
	Hispanic (0)					Non-Hispanic (1)				
	<i>N</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis	<i>N</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
VPK Assessment – AP3										
Print Knowledge	218	9.11	3.26	-0.91	-0.50	357	9.30	3.10	-0.97	-0.25
Phonological Awareness	218	8.79	3.63	-0.27	-0.99	357	9.68	3.78	-0.55	-0.58
Mathematics	218	11.41	4.23	-0.39	-0.84	357	12.46	4.17	-0.76	-0.04
Oral Language/Vocabulary	218	17.13	3.79	-0.94	0.66	357	18.91	3.15	-1.67	5.29
Teaching Strategies GOLD - Spring										
Social - Emotional	194	57.42	7.67	-0.22	0.21	339	57.24	8.43	-0.18	0.03
Cognitive	194	60.09	8.50	-0.16	1.23	339	59.79	8.99	0.23	0.51
Literacy	193	59.79	14.02	-0.11	-0.17	337	62.75	14.19	-0.10	0.02
Mathematics	194	31.92	6.14	-0.45	0.60	337	33.32	6.34	-0.45	0.30

Table 11

Samples Sizes for Work Sampling System (WSS) Categorical Variables by Main Effects for Cohort 1

	<i>N</i>	Gender		Ethnicity	
		Female (0)	Male (1)	Hispanic (0)	Non-Hispanic (1)
		<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>
FLKRS – Work Sampling System (WSS)					
Not Yet (Not Ready)	13	3 (1.4%)	10 (4.6%)	4 (2.6%)	9 (3.1%)
In Process (Ready)	217	103 (46.6%)	114 (52.3%)	76 (50.0%)	141 (49.1%)
Proficient (Ready)	209	115 (52.0%)	94 (43.1%)	72 (47.4%)	137 (47.8%)
Total	439	221	218	152	287

Table 12

Descriptive Statistics: Full Sample for Cohort 2

Variable	<i>N</i>	<i>M</i>	<i>SD</i>	Range	Skewness	Kurtosis
VPK Assessment – AP3						
Print Knowledge	556	9.05	3.34	0-12	-0.89	-0.51
Phonological Awareness	556	9.76	3.40	0-14	-0.65	-0.46
Mathematics	556	11.88	4.20	0-18	-0.51	-0.56
Oral Language/Vocabulary	556	18.70	6.54	6-23	-1.06	0.75
Teaching Strategies GOLD – Spring						
Social – Emotional	500	58.50	9.63	0-81	-1.43	5.09
Cognitive	500	60.37	9.13	32-90	0.08	0.28
Literacy	492	62.71	13.76	14-92	-0.20	-0.10
Mathematics	498	39.32	7.47	17-56	-0.17	-0.20

Note. The potential range of each subscale of the VPK Assessment is consistent with the actual range listed within the table.

Table 13

Descriptive Statistics: Gender Sample for Cohort 2

Variable	Gender									
	Female (0)					Male (1)				
	<i>N</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis	<i>N</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
VPK Assessment – AP3										
Print Knowledge	286	9.27	3.26	-1.06	-0.04	270	8.81	3.42	-0.72	-0.87
Phonological Awareness	286	9.75	3.39	-0.65	-0.56	270	9.77	3.42	-0.65	-0.34
Mathematics	286	12.05	4.17	-0.55	-0.53	270	11.71	4.24	-0.47	-0.58
Oral Language/Vocabulary	286	18.80	3.47	-1.00	0.63	270	18.59	3.62	-1.12	-1.18
Teaching Strategies GOLD - Spring										
Social - Emotional	262	59.86	8.98	-1.74	7.84	238	57.01	10.11	-1.19	3.53
Cognitive	262	61.29	8.90	0.06	0.23	238	59.34	9.28	0.12	0.39
Literacy	257	64.50	13.47	-0.29	0.58	235	60.75	13.82	-0.10	-0.14
Mathematics	262	39.98	7.32	-0.27	-0.13	236	38.58	7.58	-0.04	-0.18

Table 14

Descriptive Statistics: Ethnicity Sample for Cohort 2

Variable	Ethnicity									
	Hispanic (0)					Non-Hispanic (1)				
	<i>N</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis	<i>N</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
VPK Assessment – AP3										
Print Knowledge	183	8.92	3.39	-0.82	-0.72	373	9.11	3.33	-0.92	-0.39
Phonological Awareness	183	8.83	3.60	-0.37	-0.87	373	10.22	3.21	-0.78	-0.12
Mathematics	183	11.27	4.37	-0.46	-0.60	373	12.18	4.10	-0.52	-0.58
Oral Language/Vocabulary	183	17.80	3.96	-0.82	-0.05	373	19.14	3.24	-1.15	1.22
Teaching Strategies GOLD - Spring										
Social - Emotional	163	58.49	10.58	-2.40	10.08	337	58.51	9.16	-0.72	0.73
Cognitive	163	59.75	9.04	0.08	0.65	337	60.66	9.17	0.07	0.14
Literacy	161	61.40	14.16	-0.05	-0.50	331	63.35	13.53	-0.28	0.19
Mathematics	162	38.28	7.75	-0.06	-0.31	336	39.82	7.29	-0.20	-0.11

Table 15

Samples Sizes for Work Sampling System (WSS) Categorical Variables by Main Effects for Cohort 2

	<i>N</i>	Gender		Ethnicity	
		Female (0)	Male (1)	Hispanic (0)	Non-Hispanic (1)
		<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>
FLKRS – Work Sampling System (WSS)					
Not Yet (Not Ready)	19	9 (3.5%)	10 (4.2%)	10 (6.2%)	9 (2.7%)
In Process (Ready)	217	102 (40.0%)	115 (47.9%)	67 (41.4%)	150 (45.0%)
Proficient (Ready)	259	144 (56.5%)	115 (47.9%)	85 (52.4%)	174 (52.3%)
Total	495	255	240	162	333

Group Statistics

Effect-sizes (Cohen's *d*) were computed using standardized mean differences to compare the males and females on the VPK Assessment subscales and Teaching Strategies GOLD® subscales, and the Hispanic and non-Hispanic groups on the same subscales. Effect-sizes of 0.2 indicate a small difference; 0.5 indicate a medium difference, and 0.8 or larger as a large difference (Cohen, 1988). Tables 16 through 23 display the effect-sizes by cohort. Overall, a majority of effect-sizes across cohorts were small ($d = -0.31 - 0.32$). A medium difference between the Hispanic and non-Hispanic children was noticed within cohort 1 on the VPK Assessment Oral Language/Vocabulary subscale ($d = -0.52$). This difference indicated that the non-Hispanic children scored nearly a half a standard deviation higher than their Hispanic peers on the Oral Language/Vocabulary subscale.

Table 16

Descriptive Statistics and Effect Sizes by Gender Based on Main Effects - VPK Assessment subscales – Cohort 1

Subtest	Female			Male			<i>d</i>	95% Confidence Interval for <i>d</i>	
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>		Lower	Upper
	Print Knowledge	274	9.56	2.89	301	8.92		3.37	0.20
Phonological Awareness	274	9.51	3.40	301	9.18	3.59	0.09	-0.07	0.26
Mathematics	274	12.42	3.84	301	11.74	4.52	0.16	-0.00	0.32
Oral Language/ Vocabulary	274	18.42	3.37	301	18.07	3.63	0.10	-0.06	0.26

Table 17

Descriptive Statistics and Effect Sizes by Ethnicity Based on Main Effects – VPK Assessment subscales – Cohort 1

Subtest	Hispanic			Non-Hispanic			<i>d</i>	95% Confidence Interval for <i>d</i>	
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>		Lower	Upper
	Print Knowledge	218	9.11	3.26	357	9.30		3.10	-0.06
Phonological Awareness	218	8.79	3.63	357	9.68	3.78	-0.24	-0.41	-0.07
Mathematics	218	11.41	4.23	357	12.46	4.17	-0.31	-0.48	-0.15
Oral Language/ Vocabulary	218	17.13	3.79	357	18.91	3.15	-0.52	-0.70	-0.35

Table 18

Descriptive Statistics and Effect Sizes by Gender Based on Main Effects - Teaching Strategies GOLD subscales – Cohort 1

Subtest	Female			Male			<i>d</i>	95% Confidence Interval for <i>d</i>	
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>		Lower	Upper
	Social-Emotional	254	58.69	7.94	256	56.13		8.21	0.32
Cognitive	254	60.76	8.35	256	59.18	9.24	0.18	0.01	0.35
Literacy	252	63.62	13.22	256	60.13	14.98	0.25	0.07	0.42
Mathematics	253	33.53	5.67	256	32.37	6.71	0.19	0.01	0.36

Table 19

Descriptive Statistics and Effect Sizes by Ethnicity Based on Main Effects - Teaching Strategies GOLD subscales – Cohort 1

Subtest	Hispanic			Non-Hispanic			<i>d</i>	95% Confidence Interval for <i>d</i>	
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>		Lower	Upper
	Social-Emotional	187	57.41	7.73	323	57.40		8.42	0.00
Cognitive	187	60.06	8.57	323	59.91	9.00	0.02	-0.16	0.20
Literacy	186	59.72	14.13	322	63.09	14.16	-0.24	-0.42	-0.06
Mathematics	187	31.96	6.07	322	33.52	6.26	-0.25	-0.43	-0.07

Table 20

Descriptive Statistics and Effect Sizes by Gender Based on Main Effects - VPK Assessment subscales – Cohort 2

Subtest	Female			Male			<i>d</i>	95% Confidence Interval for <i>d</i>	
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>		Lower	Upper
	Print Knowledge	286	9.27	3.26	270	8.81		3.42	0.14
Phonological Awareness	286	9.75	3.39	270	9.77	3.42	-0.01	-0.17	0.16
Mathematics	286	12.05	4.17	270	11.71	4.24	0.08	-0.09	0.25
Oral Language/ Vocabulary	286	18.80	3.47	270	18.59	3.62	0.06	-0.11	0.23

Table 21

Descriptive Statistics and Effect Sizes by Ethnicity Based on Main Effects – VPK Assessment subscales – Cohort 2

Subtest	Hispanic			Non-Hispanic			<i>d</i>	95% Confidence Interval for <i>d</i>	
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>		Lower	Upper
	Print Knowledge	183	8.92	3.39	373	9.11		3.33	-0.06
Phonological Awareness	183	8.83	3.60	373	10.22	3.21	-0.40	-0.59	-0.23
Mathematics	183	11.27	4.37	373	12.18	4.10	-0.22	-0.39	-0.04
Oral Language/ Vocabulary	183	17.80	3.96	373	19.14	3.24	-0.38	-0.56	-0.21

Table 22

Descriptive Statistics and Effect Sizes by Gender Based on Main Effects - Teaching Strategies GOLD subscales – Cohort 2

Subtest	Female			Male			<i>d</i>	95% Confidence Interval for <i>d</i>	
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>		Lower	Upper
	Social-Emotional	262	59.86	8.98	238	57.01		10.11	0.30
Cognitive	262	61.29	8.90	238	59.34	9.29	0.21	0.04	0.39
Literacy	257	64.50	13.47	235	60.75	13.82	0.28	0.10	0.45
Mathematics	262	39.98	7.32	236	38.58	7.58	0.19	0.01	0.36

Table 23

Descriptive Statistics and Effect Sizes by Ethnicity Based on Main Effects - Teaching Strategies GOLD subscales – Cohort 2

Subtest	Hispanic			Non-Hispanic			<i>d</i>	95% Confidence Interval for <i>d</i>	
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>		Lower	Upper
	Social-Emotional	163	58.46	10.58	337	58.51		9.16	0.00
Cognitive	163	59.75	9.04	337	60.66	9.17	- 0.03	- 0.21	0.16
Literacy	161	61.40	14.16	331	63.35	13.53	- 0.14	- 0.33	0.05
Mathematics	162	38.28	7.75	336	39.82	7.29	- 0.21	- 0.39	- 0.02

Model Results

Hosmer, Lemeshow, and Sturdivant (2013) state that although ordinal logistic regression does not follow typical normality assumptions, it is important to determine that there is little or

no multicollinearity among the independent variables of models. In other words, independent variables should not be too highly correlated with each other. In order to screen for this assumption, correlation matrices for each cohort were created for the VPK Assessment subscales; Teaching Strategies GOLD® subscales; and VPK Assessment Subscales and Teaching Strategies GOLD® subscales. Pearson's *rs* were examined within each matrix with small to moderate correlations noted. However, within both Cohort 1 and Cohort 2, strong correlations were found between the Teaching Strategies GOLD® Literacy and Mathematics subscales ($r = .88$ and $r = .87$, Cohort 1 and Cohort 2, respectively).

As evidence of potential multicollinearity was present between the Teaching Strategies GOLD® Literacy and Mathematics subscales, I then examined how the models looked with both variables in the model and then with only one subscale (either Literacy or Mathematics). Specifically, for both Cohort 1 and 2, I ran each model without the specific subscale that was indicating multicollinearity to determine if the model would change. For the model which examined the potential interaction between gender and Literacy (gender X Literacy), I removed the mathematics subscale from the model. For the model that examined the interaction between gender and Mathematics (gender X Mathematics), I removed literacy from the model. I continued this same process for the models that examined interactions between ethnicity and the literacy and mathematics subscales independently. The outcome of this process was that by taking the mathematics subscales out of the models that were examining the interactions between the variables of gender and ethnicity, no differences were present. The interactions continued to be not statistically significant. The same results were noted when removing the literacy subscales from the models that focused on the interactions between the control variables of gender and ethnicity with the mathematics subscales. Therefore, I determined that although

there was a strong relationship between the Teaching Strategies GOLD® Literacy and Mathematics subscales, the degree of multicollinearity did not affect the results.

As the dependent variable of kindergarten readiness was categorical in nature, ordinal logistic regression analyses were conducted to determine the predictive ability of the VPK Assessments and Teaching Strategies GOLD® on kindergarten readiness. The goal of these ordinal logistic regression models was to examine the relation between kindergarten readiness (outcome) and several assessment measures (independent variables). Ordinal logistic regression was to be used to determine the probability of children being not yet ready for kindergarten (category 0), in process of being ready for kindergarten (category 1) or proficient in kindergarten readiness (category 2).

In order to address the research questions for this study, a series of models were run. I first looked at each assessment measure subscale itself in a model with gender and ethnicity as control variables. The same process was used to examine the relationships between the control variables (gender and ethnicity), main effects of VPK Assessment subscales, and the interaction between ethnicity and each individual VPK Assessment subscale. The same modeling process was followed when examining the Teaching Strategies GOLD® subscales.

As this secondary analysis consisted of a sample of children nested in classrooms that were nested within school sites, *Mplus* (Muthén & Muthén, 1998-2017) was utilized to adjust the standard errors presented within the models. Data for the multilevel modeling of children nested within classrooms are reported as the regression coefficient (β), standard error of the regression coefficient ($SE \beta$), regression coefficient/standard error of the regression coefficient ($\beta / SE \beta$) and *p*-value. To evaluate the sensitivity of the results to the nested data structure, all models were re-run not taking into account the nested data structure. The Huber-White sandwich

estimator was utilized in each model to estimate the variance of the Maximum Likelihood Ratios if the underlying models were incorrect (Freedman, 2005). This sandwich estimator was added into the *Mplus* model syntax with the command `Analysis: type = complex`. An example of the complete *Mplus* syntax for a cohort 1 model examining the relationship between the VPK Assessment subscale Print Awareness and gender can be found in Appendix C. These tables are presented in Appendix D. Due to each cohort having a generous sample size, I set more stringent alpha levels (α) at .01 and 99% confidence interval estimate for the odds ratio.

This quantitative study addressed the following research questions:

RQ1. What is the relationship between the VPK Assessment subscales (Print Knowledge, Phonological Awareness, Mathematical Concepts, Oral Language/Vocabulary) and Teaching Strategies GOLD® (Social-Emotional, Cognitive, Literacy and Mathematics) and kindergarten readiness (Not Ready, In Process, and Proficient)?

RQ2. To what extent are the relationships between the VPK Assessment subscales (Print Knowledge, Phonological Awareness, Mathematical Concepts, and Oral Language/Vocabulary) and kindergarten readiness (Not Ready, In Process, and Proficient) moderated by gender (female vs. male)?

RQ3. To what extent are the relationships between VPK Assessment subscales (Print Knowledge, Phonological Awareness, Mathematical Concepts, and Oral Language/Vocabulary) and kindergarten readiness (Not Ready, In Process, and Proficient) moderated by ethnicity (Hispanic vs. Non-Hispanic)?

RQ4. To what extent are the relationships between Teaching Strategies GOLD® subscales (Social-Emotional, Cognitive, Literacy, and Mathematics) and kindergarten readiness (Not Ready, In Process, and Proficient) moderated by gender (female vs. male)?

RQ5. To what extent are the relationships between Teaching Strategies GOLD® subscales (Social-Emotional, Cognitive, Literacy, and Mathematics) and kindergarten readiness (Not Ready, In Process, and Proficient) moderated by ethnicity (Hispanic vs. Non-Hispanic)?

Models examining scales independently.

Data from both cohorts were used to evaluate the predictive validity and differential predictive validity of each assessment measure subscale independently. I first looked at each assessment measure subscale by itself in a model with gender and ethnicity as control variables. In order to conclude that there was a robust predictive relationship, the relationship had to be evident within both cohorts. Table 24 illustrates the results of examining the predictive and differential relationships of each assessment measure subscale independently. Table 24 shows that each VPK Assessment subscale had a robust, positive predictive relationship to kindergarten readiness as measured by the WSS. The Print Knowledge and Mathematics subscales had the strongest predictive relationships across both cohorts. When exploring the differential validity relationships between the VPK Assessment subscales and gender (female vs. male) there was a lack of robust interaction effects. These results indicated no evidence of predictive bias in the manner that each VPK Assessment subscale was functioning similarly in predicting kindergarten readiness for the female and male groups. In summary, each subscale of the VPK Assessment was working similarly for females or males. As with gender, when examining the interactions between each VPK Assessment subscale and ethnicity (Hispanic vs. Non-Hispanic) a lack of robust interaction effects was present. The results indicated a lack of predictive bias in that each VPK Assessment subscale was functioning similarly in predicting kindergarten readiness for the Hispanic and Non-Hispanic groups. These data suggested that each VPK Assessment subscale was working similarly for Hispanic and Non-Hispanic children.

Table 24

Summary of Predictor and Interaction Relationships by Each Assessment Measure Subscale

Measure	Cohort 1			Cohort 2			Conclusions		
	Only Predictor	Predictor X Gender	Predictor X Ethnicity	Only Predictor	Predictor X Gender	Predictor X Ethnicity	Conclusion About Gender Interaction	Conclusion About Ethnicity Interaction	Conclusion About Predictor
VPK Assessment									
Print Knowledge	0.09*	0.04	0.03	0.08*	0.00	- 0.02	Not Robust	Not Robust	Robust
Phonological Awareness	0.06*	0.02	0.02	0.05*	- 0.01	- 0.05	Not Robust	Not Robust	Robust
Mathematics	0.07*	0.05	0.02	0.08*	- 0.01	- 0.06*	Not Robust	Not Robust	Robust
Oral Language/Vocabulary	0.06*	0.02	- 0.01	0.07*	- 0.02	- 0.03	Not Robust	Not Robust	Robust
Teaching Strategies Gold									
Social Emotional	0.03*	0.01	0.01	0.02*	0.01	- 0.04*	Not Robust	Not Robust	Robust
Cognitive Literacy	0.04*	0.00	- 0.02	0.02	- 0.00	- 0.03	Not Robust	Not Robust	Not Robust
Mathematics	0.03*	0.01	0.00	0.02*	0.00	- 0.02	Not Robust	Not Robust	Robust
	0.05*	0.02	0.00	0.04*	0.01	- 0.63	Not Robust	Not Robust	Robust

Note: * $p \leq .01$; when looking at the three models; (1) one assessment measure; (2) one assessment measure and the interaction of the assessment measure and gender; and (3) one assessment measure and the interaction of the assessment measure and ethnicity.

Additionally, Table 24 displays that each Teaching Strategies GOLD® subscale with the exception of Cognitive has a robust predictive relationship for kindergarten readiness as measured by the WSS. The Mathematics and Literacy subscales had the strongest predictive relationships across both cohorts. When exploring the differential validity relationships for the Teaching Strategies GOLD® subscales by gender (female vs. male) there was a lack of robust interaction effects. These results indicated a lack of predictive bias in that each Teaching Strategies GOLD® subscale was functioning similarly in predicting kindergarten readiness for the female and male groups. As with gender, when examining the interactions between each Teaching Strategies GOLD® subscale and ethnicity (Hispanic vs. Non-Hispanic), there was a lack of robust interaction effects. The results indicated a lack of predictive bias in that each Teaching Strategies GOLD® subscale was functioning similarly in predicting kindergarten readiness for the Hispanic and Non-Hispanic groups.

When examining test-criterion relationships, evidence about relations to other variables can be used to investigate questions of differential prediction for subgroups. If the “relation of test scores to a relevant criterion variable differs from one subgroup to another may imply that the meaning of the scores is not the same for members of the different groups, perhaps due to construct underrepresentation or to construct-irrelevant sources of variance” (AERA, APA, NCME, 2014, p. 18). These results, which showed no evidence of subgroup differences with each subscale of the VPK Assessment and Teaching Strategies GOLD®, provide one source of evidence of the fairness of the measures.

Models with all subscales included.

Table 25 presents the results of predictive and differential predictive validity relationships of kindergarten readiness, as measured by the WSS, when: (a) all subscales of the

VPK Assessment were added to the models and analyzed simultaneously, and (b) all Teaching Strategies GOLD® subscales were added to the models and analyzed simultaneously. Overall, no predictive relationships were observed for both the VPK Assessment and Teaching Strategies GOLD® when all of the subscales of each assessment measure were analyzed simultaneously within their respective models (i.e., VPK Assessments were analyzed separately from Teaching Strategies GOLD® subscales).

The results of a decreasing predictive relationships within the models when all of the assessment subscales were analyzed simultaneously was not surprising as the more scales added to a model, the effect of any subscale by itself drops off. However, it is interesting that by simultaneously analyzing all to the subscales together within the model, there was a complete lack of predictive validity of both the VPK Assessment and Teaching Strategies GOLD® assessment measures.

When exploring differential predictive validity for the VPK Assessment and Teaching Strategies GOLD® subscales by gender (female vs. male), there were no robust interaction effects for both assessment measures. The VPK Assessment and Teaching Strategies GOLD® were predicting similarly for females and males. As with gender, when examining the differential predictive validity for the VPK Assessment and Teaching Strategies GOLD® subscales by ethnicity (Hispanic vs. Non-Hispanic), there were no robust interaction effects for either measure. These data suggested that the VPK Assessment and Teaching Strategies GOLD® subscales were working similarly for Hispanic and Non-Hispanic children.

Table 25

Summary of Predictor and Interaction Relationship by Assessment Measures with All Subscales Included

Measure	Cohort 1			Cohort 2			Conclusions		
	Predictor	Predictor X Gender	Predictor X Ethnicity	Predictor	Predictor X Gender	Predictor X Ethnicity	Conclusion About Gender Interaction	Conclusion About Ethnicity Interaction	Conclusion About Predictor
VPK Assessment									
Print Knowledge	0.01	0.04	0.03	0.02	0.01	- 0.06	Not Robust	Not Robust	Not Robust
Phonological Awareness	0.01	0.01	0.01	- 0.03	0.01	- 0.04	Not Robust	Not Robust	Not Robust
Mathematics	0.02	0.05	0.02	0.06*	- 0.01	- 0.06	Not Robust	Not Robust	Not Robust
Oral Language/Vocabulary	- 0.01	0.01	- 0.02	0.04	- 0.01	- 0.02	Not Robust	Not Robust	Not Robust
Teaching Strategies Gold									
Social Emotional	0.01	0.01	0.02	0.01	0.01	- 0.03	Not Robust	Not Robust	Not Robust
Cognitive	0.01	0.00	0.00	0.02	- 0.01	- 0.03	Not Robust	Not Robust	Not Robust
Literacy	0.02	0.01	0.00	- 0.02	0.00	- 0.02	Not Robust	Not Robust	Not Robust
Mathematics	- 0.01	0.02	0.00	0.02	0.00	- 0.04	Not Robust	Not Robust	Not Robust

Note: Nine models were run: Model 1 consisted of the control variables plus all assessment measure subscales; Models 2 – 5 consisted of the control variables plus all subscales plus one subscale x gender interaction; and Models 6-9 consisted of the control variables plus all subscales plus one subscale x ethnicity interaction.

Chapter Summary

The predictive relationships of the VPK Assessment and Teaching Strategies GOLD® were examined through the use of ordinal logistic regression models. In order to examine the robustness of the predictive relationships, analyses were conducted by examining the predictive relationships of the subscales independently and then when all subscales of the assessment measure were simultaneously included in the models. Models included the predictors of gender (Female = 0, Male = 1) and ethnicity (Hispanic = 0, Non-Hispanic = 1). Interactions of gender and each of the assessment measures and ethnicity each of the assessment measures were also examined. The operational definition of a robust predictive relationship was for a statistically significant relationship for both cohorts ($p \leq .01$). Each of the VPK Assessment subscales when examined independently (Print Knowledge, Phonological Awareness, Mathematics and Oral Language/Vocabulary) had a robust predictive relationship with the level of kindergarten readiness on the WSS. The following subscales of Teaching Strategies GOLD® exhibited robust predictive relationships with the level of kindergarten readiness on the WSS (Social Emotional, Literacy and Mathematics). The Cognitive subscale was the only scale that did not produce a robust predictive relationship with the WSS. When examining the interactions between the subscales independently for both the VPK Assessment and Teaching Strategies GOLD® and each of the demographic variables (i.e., gender and ethnicity), there were no robust interaction effects by gender or ethnicity.

Models were then created and examined for the predictive and differential predictive validity when all subscales of the VPK Assessment and Teaching Strategies GOLD® were examined simultaneously within the models. Overall, no robust predictive relations were observed for both the VPK Assessment and Teaching Strategies GOLD®.

CHAPTER FIVE

DISCUSSION

Determining school readiness at kindergarten entry using assessments has become more universal across the United States. With the emphasis on readiness, stakeholders have sought to determine if different assessments in ECE programs predict readiness. Within the state of Florida, state education agencies and policy makers have focused specifically on the predictive validity of the state-mandated, standardized VPK Assessment, while the Office of Head Start has focused on more authentic assessments, such as the Teaching Strategies GOLD® assessment, to predict readiness to start school for children participating in ECE programs. As emphasized in the *Standards for Educational and Psychological Testing*, assessment instruments that have been validated for their intended purposes have the potential to provide substantial benefits for test takers and test users (AERA, APA, NCME, 2014). These benefits include better decisions about individuals and programs than would result without their use and can also provide a route to broader and more equitable access to education and employment. However, the improper use of tests has the ability to unintentionally cause considerable harm to test takers and other parties affected by test-based decisions.

The purpose of this quantitative, correlational study was to understand how a standards-based measure, Florida VPK Assessments (FLDOE, 2011) and a curriculum-embedded measure, Teaching Strategies GOLD®, administered in the preschool year within a blended Head Start/VPK program predict kindergarten readiness within the state of Florida. The study focused on determining if predictive relationships exist between the VPK Assessments and Teaching Strategies GOLD® scores and the criterion variable (dependent variable) of school readiness at

kindergarten entry, as measured by the Florida Kindergarten Readiness Screener (FLKRS) Work Sampling System™ (WSS). The dependent variable of school readiness from the WSS was coded *Not Ready*, *In Process*, and *Proficient*. Additionally, the study examined whether selected child characteristics of gender and ethnicity (Hispanic vs. Non-Hispanic) moderated the predictive relationships between the VPK Assessments and Teaching Strategies GOLD® measures and the dependent variable of school readiness at kindergarten entry.

Chapter Five presents a summary of the research study and interpretations of the specific findings. Specifically, the chapter summarizes the levels of predictive and differential validity of the VPK Assessment subscales and Teaching Strategies GOLD® scales with the outcome of school readiness at kindergarten entry with the FLKRS WSS™. Limitations of the study are provided as well as discussion for future research.

Review of Sample and Methods

The research sample from the secondary analysis consisted of two cohorts of preschool-aged children enrolled within a blended funding model of the federal income-based Head Start program and the universal state voluntary prekindergarten program (Florida VPK). Each cohort consisted of children who were nested within 39 classrooms within 22 elementary schools. Cohort 1 children were enrolled within the blended Head Start/VPK classrooms during the 2014-2015 academic year, while Cohort 2 children were enrolled during the 2015-2016 academic year.

Data from the VPK Assessment and Teaching Strategies GOLD® assessment were collected at the end of May of each cohort year (2015 and 2016, respectively), while the FLKRS Work Sampling System WSS® data were collected by kindergarten teachers at the end of September (2015 and 2016, respectively). The VPK Assessment variables include Print Knowledge, Phonological Awareness, Mathematical Concepts, and Oral Language/Vocabulary.

The Teaching Strategies GOLD® assessment variables include Social-Emotional, General Cognition (focusing on approaches to learning), Literacy, and Mathematics.

Examination of descriptive statistics of the VPK Assessment subscales for both cohorts indicated a majority of children scored at the mid-range on all VPK Assessment subscales. Females and males had similar achievement on all VPK Assessment subscales with the exception of Print Knowledge with females scoring slightly higher than males. When looking at the descriptive data for the VPK Assessment subscales with regard to ethnicity, a difference in achievement across subscales was indicated with more Non-Hispanic children in both cohorts scoring in the mid- to high-range than their Hispanic peers across all subscales.

The examination of the descriptive statistics of the Teaching Strategies GOLD® subscales for both cohorts indicated that a majority of children scored in the mid- to high-range across all subscales. Females and males had similar achievement on all subscales with the exception of the Social-Emotional and Literacy subscales, in which the females scored at a higher level. Descriptive data focused on group differences by ethnicity among the Teaching Strategies GOLD® subscales showed across both cohorts that Non-Hispanic and Hispanic children had similar achievement on all subscales apart from the Literacy subscale. Non-Hispanic children scored higher on literacy.

Kindergarten readiness was determined utilizing the categorical composite scores of the FLKRS WSS® of *Not Yet* (0), *In Process* (1), and *Proficient* (2). Within Cohort One, the distribution of the sample related to kindergarten readiness on the WSS® achieving the levels of *Not Yet* (3.0%), *In Process* (49.4%), and *Proficient* (47.6%). Cohort Two had a similar distribution of kindergarten readiness on the WSS® with children scoring within the composite scores of *Not Yet* (3.8%), *In Process* (43.9%), and *Proficient* (52.3%). Both cohorts had 96% of

the sample scoring at the *In Process* or *Proficient* range on the WSS® kindergarten readiness assessment. The descriptive data for the FLKRS WSS® followed a similar pattern of the independent assessment measures across both cohorts. Both males and Hispanic children had a larger rate (nearly double and triple, respectively) of children achieving a kindergarten readiness rate of *Not Yet* compared to their female and Non-Hispanic peers.

Multilevel ordinal regression analyses of children nested within classrooms were used to evaluate the predictive relationship between the standards-based measure of the VPK Assessment and curriculum-embedded measure of Teaching Strategies GOLD® and the outcome of kindergarten readiness (FLKRS WSS). Additional ordinal regression analyses were used to determine if the previously stated relationships were moderated by child gender or ethnicity (Hispanic vs. Non-Hispanic).

Findings

The five research questions within this study corresponded to two validity issues: (1) did the assessment measures predict kindergarten readiness; and (2) to what extent was the predictive relationship different amongst groups (gender and ethnicity). The rationale for answering the questions is supported by the Standard 7.12 of the *Standards for Educational and Psychological Testing*, which states that when test scores are used to make predictions about future behavior, the evidence supporting those predictions should be provided to the test user (AERA, APA, NCME, 2014). Additionally, the “relation of test scores to a relevant criterion variable differs from one subgroup to another may imply that the meaning of the scores is not the same for members of the different groups, perhaps due to construct underrepresentation or to construct-irrelevant sources of variance” (AERA, APA, NCME, 2014, p. 18). Finally, the authors also state within Standard 3.7 that when criterion-related validity evidence is used for

test score-based predictions of future performance, test users need to be responsible for evaluating the possibility of differential prediction for relevant subgroups when there is prior evidence or theory suggesting differential prediction (AERA, APA, NCME). Based on information gathered from current literature, this study focused on the potential for differential prediction within gender subgroups (female vs. male) and ethnicity (Non-Hispanic vs. Hispanic). Both standards were foundational in the creation of the predictive validity and differential predictive questions as validity information was evaluated for the two measures (VPK Assessment, Teaching Strategies GOLD®) utilized within the study

Models examining scales independently.

Florida VPK Assessment Measures. Data from two cohorts were used to evaluate the predictive validity and differential predictive validity of each subscale. Based on a series of models in which each scale was examined separately, with gender and ethnicity as control variables, each VPK Assessment subscale had a robust, positive predictive relationship to kindergarten readiness. In order to conclude that there was a robust predictive relationship, the statistically-significant relationship had to be evident within both cohorts. The Print Knowledge and Mathematics subscales had the strongest predictive relationships across both cohorts. When exploring the differential validity relationships between the VPK Assessment subscales and gender (female vs. male) there was a lack of robust interactions observed. These results indicated a lack of predictive bias suggesting that there was insufficient evidence to conclude that each VPK Assessment subscale was functioning differently in predicting kindergarten readiness for the female and male groups. As with gender, when examining the interactions between each VPK Assessment subscale and ethnicity (Hispanic vs. Non-Hispanic) a lack of robust interactions was present. The results indicated a lack of predictive bias suggesting that

there was insufficient evidence to conclude that each VPK Assessment subscale was functioning differently in predicting kindergarten readiness for the Hispanic and Non-Hispanic groups.

Teaching Strategies GOLD® Assessment. Each Teaching Strategies GOLD® subscale, with the exception of the Cognitive scale, had a robust, positive predictive relationship for kindergarten readiness. The Mathematics and Literacy subscales had the strongest predictive relationships across both cohorts. When exploring the differential validity relationships between the Teaching Strategies GOLD® subscales and gender (female vs. male) there was a lack of robust interactions observed. These results indicated a lack of predictive bias suggesting that there was insufficient evidence to conclude that each Teaching Strategies GOLD® subscale was functioning differently in predicting kindergarten readiness for the female and male groups. As with gender, when examining the interactions between each Teaching Strategies GOLD® subscale and ethnicity (Hispanic vs. Non-Hispanic) a lack of robust interactions was present. The results indicated a lack of predictive bias suggesting that there was insufficient evidence to conclude that each Teaching Strategies GOLD® subscale was functioning differently in predicting kindergarten readiness for the Hispanic and Non-Hispanic groups.

When examining test-criterion relationships, evidence about relations to other variables can be used to investigate questions of differential prediction for subgroups. If the “relation of test scores to a relevant criterion variable differs from one subgroup to another [this] may imply that the meaning of the scores is not the same for members of the different groups, perhaps due to construct underrepresentation or to construct-irrelevant sources of variance.” (AERA, APA, NCME, 2014, p. 18). These results, which did not show any evidence of subgroup differences for each subscale of the VPK Assessment and Teaching Strategies GOLD®, did not identify any concerns about the fairness of the assessments in predicting kindergarten readiness.

The presence of a robust predictive relationship between each of the VPK Assessment subscales and kindergarten readiness supported the results of the VPK Assessment subscales with the previous FLKRS curriculum-embedded measure, the Early Childhood Observation Screener (ECHOS), which showed moderate validity coefficients ($r = .20$ to $.32$). The two cohorts of this study (2014-2015 and 2015-2016) had VPK Assessment subscales and WSS results that also showed moderate validity coefficients ($r = .17$ to $.25$ and $r = .15$ to $.28$, respectively). The VPK Assessment subscales with the strongest validity coefficients with the WSS were Mathematics ($r = .25$ and $r = .28$, respectively) and Print Knowledge ($r = .24$ and $r = .22$, respectively).

Additionally, the authors of the VPK Assessment measures (FLDOE, 2011) stated that the goal of the measure was to provide stakeholders with a means of identifying children who are not on a trajectory of success to exhibit kindergarten readiness in terms of their developing reading-related and mathematics skills during the children's VPK experience. The robust predictive relationships between the VPK Assessment subscales were supportive of the authors' claim of item development of the measures involved creating sets of items that were mapped onto domains of early literacy and mathematics that were included in *Florida's Early Learning and Developmental Standards for Four-Year-Olds* (FLDOE, 2011b).

Russo, Williford, Markowitz, Vitello and Bassok (2019) state that to date there is no research that has examined the validity of Teaching Strategies GOLD® longitudinally. As such, the selected subscales of Teaching Strategies GOLD® were not examined with the previous FLKRS curriculum-embedded measure, the Early Childhood Observation Screener. However, validity coefficients for Teaching Strategies GOLD® subscales and the WSS were reviewed for the two cohorts of this study (2014-2015 and 2015-2016). The Teaching Strategies GOLD®

subscales and WSS validity coefficients were moderate in size ($r = .21$ to $.31$ and $r = .17$ to $.27$, respectively). The Teaching Strategies GOLD® subscales with the strongest validity coefficients were Literacy ($r = .31$ and $r = .27$, respectively) and Mathematics ($r = .27$ and $r = .23$, respectively). The presence of a robust predictive relationship between each of the Teaching Strategies GOLD® subscales were supportive of the authors' claim of a crosswalk between the sets of items that were mapped onto domains of social-emotional, early literacy, and mathematics with the *Florida's Early Learning and Developmental Standards for Four-Year-Olds* (FLDOE, 2011b). These predictive relationships of kindergarten readiness were also similar to the results of the content validity studies conducted by the authors of Teaching Strategies GOLD® with the Literacy subscale exhibiting the strongest association among the subscales.

Models with all subscales included.

Based on the literature review for this study, no evidence was found of studies with multiple logistic models in which all VPK Assessment subscales were simultaneously included to predict kindergarten readiness (measured using the WSS). Similarly, no studies were found that used multiple logistic models in which all of the Teaching Strategies GOLD® subscales were simultaneously included to predict kindergarten readiness (measured using the WSS). Overall, no predictive relationships were observed for both the VPK Assessment and Teaching Strategies GOLD® when all of the subscales of each assessment measure were simultaneously considered within the model. The results of decreasing predictive relationships within the models simultaneously including all of the assessment measures' subscales were not surprising as the more scales that are correlated are added to a model, the effect of any subscale by itself drops off. Within a logistic regression model, the predictor variables are often related to each

other and compete with each other in explaining the outcome. For the VPK Assessment, the correlations between the subscales ranged from .33 to .68 with an average correlation of .53 for Cohort 1 and .54 for Cohort 2. For the Teaching Strategies GOLD® subscales, the correlations between the subscales ranged from .57 to .88 with an average correlation of .73 for Cohort 1 and .71 for Cohort 2. Therefore, by including all of the VPK Assessment and Teaching Strategies GOLD® subscales together to determine each assessment measure's predictive validity with the WSS, a natural reduction in predictability occurred. By including all of the subscales together within each assessment measure's model, a complete lack of predictive validity of both the VPK Assessment and Teaching Strategies GOLD® assessment measures was evident as the moderate to strongly correlated subscales were competing with each other within the model, in turn reducing the predictive validity.

When exploring the differential validity relationships between the VPK Assessment and Teaching Strategies GOLD® subscales and gender (female vs. male) there was a lack of robust relationships observed with both assessment measures. The VPK Assessment and Teaching Strategies GOLD® were working similarly for females or males. As with gender, when examining the interactions between the VPK Assessment and Teaching Strategies GOLD® subscales and ethnicity (Hispanic vs. Non-Hispanic) a lack of robust relationships was present. This data suggested that the VPK Assessment and Teaching Strategies GOLD® subscales were working similarly for Hispanic and Non-Hispanic children.

Limitations of the Study

The authors of the *Standards for Educational and Psychological Testing* (2014) state that validation is the joint responsibility of the test developer and test user. Therefore, when a test user proposes an interpretation or use of test scores that differs from those supported by the test

developer, the responsibility for providing validity evidence in support of that interpretation of the specified use is the responsibility of the test user. In the case of this study, the test user was the state of Florida that made the determination to use the commercially-made Work Sampling System™ as the state-mandated kindergarten readiness measure.

This predictive validity study was a secondary analysis that used a convenience sample of preschool children enrolled in a blended Head Start/state-funded voluntary prekindergarten program (VPK) from one school district in Florida. Although this study found evidence of robust predictive relationships of the independent subscales of the VPK Assessments and each subscale of Teaching Strategies GOLD® with the exception of Cognitive, as well as finding no differential predictive validity evidence across gender and ethnicity groups, the extent to which predictive validity evidence can be generalized to new situations is a function of accumulated research. In order to accumulate additional predictive and differential validation research related to the VPK Assessments and Teaching Strategies GOLD® subscales, secondary data cohorts should be gathered from similar like-sized school districts with similar demographic distributions in Florida that provided blended Head Start/VPK classes within the 2014-2015 and 2015-2016 school years.

Curriculum-embedded and authentic assessments, such as Teaching Strategies GOLD®, often require practitioners to observe children for an established period of time and use an assessment tool to determine the appropriate outcome. Standard 4.21 of the *Standards for Educational and Psychological Testing* states that test users are responsible for scoring and when scoring requires scorer judgement, the test user is responsible for providing adequate training and instruction to the scorers and for examining scorer agreement and accuracy. The test developer should document the expected level of agreement and accuracy and should provide as much

technical guidance as possible to aid test users in satisfying this standard. Within this study, Head Start/VPK teachers were used to determine the levels of each child on the specific subscales in the curriculum-embedded Teaching Strategies GOLD®. Teachers were provided with three-days of face-to-face professional development by the test developers on the appropriate manner to collect anecdotal information and appropriately rate each child on a scale of one to nine on each indicator within each subscale. Within each day of professional development, authentic opportunities with the use of video clips were provided for teachers to observe children, collect anecdotal information, and select an appropriate rating on each indicator. Teachers were provided individualized feedback from the professional development coordinators. In addition to the three days of professional development, teachers were required to participate and adequately complete four computerized training modules prior to the beginning of the 2014-2015 school year. As determined by the assessment authors, adequate completion of the online professional development course was to receive an 80% or higher on each embedded formative and summative assessments. For teachers needing additional support, Head Start Instructional Coaches provided individualized support to said teachers to ensure appropriate anecdotal recordings and subscale ratings.

Miller-Bains et al. (2017) conducted a study on the utility of Teaching Strategies GOLD® as a Kindergarten Entry Assessment (KEA). The authors found the measure demonstrated weakness in its ability to differentiate readiness skills. The authors found highly correlated subscales and intraclass correlations with the same classroom relative to the direct assessment. The authors suggested that the teachers within the study tended to rate individual children more similarly across all learning constructs despite empirical evidence of more substantial variation across domains when skills are measures via direct assessment, with

Teaching Strategies GOLD® less able to differentiate children's skills in a specific learning area within a classroom (Miller-Bains, 2017). Within the current study, Pearson's *rs* were examined within each matrix. Within both cohorts, strong correlations were found between the Teaching Strategies GOLD® Literacy and Mathematics subscales ($r = .88$ and $r = .87$, Cohort 1 and Cohort 2, respectively). As the current study also indicates levels of potential weakness in the ability to differentiate readiness skills, especially within the areas of Literacy and Mathematics, further research should be conducted to determine if additional professional development, monitoring and coaching for the assessment measure is needed in order for the Teaching Strategies GOLD® to be utilized appropriately for its intended purpose.

Unlike professional development opportunities provided to teachers for VPK Assessments and Teaching Strategies GOLD®, the state of Florida did not provide school districts with funding to provide in-depth professional development for scoring the WSS for all kindergarten teachers within public and charter schools. The Bureau of Just Read, Florida! within the Florida Department of Education provided a half-day regional professional development opportunity for school district assessment coordinators and coordinators of elementary programs. At this meeting, district coordinators were exposed to the purpose and construct of the WSS. In addition, coordinators were provided with a training powerpoint with the expectation to be shared with school-based administrators and kindergarten teachers. As the state provided districts with no funding to train union-contracted kindergarten teacher ($N \geq 140$) over the summer, this district trained one school-based administrator on how to administer the WSS. The expectation was for the school-based administrator to ensure that all kindergarten teachers within their school be trained on how to administer the WSS within the teacher planning week prior to children entering school in August. However, district assessment coordinators had

no way of monitoring the fidelity of this ‘trickle-down’ training of the WSS at 50 elementary schools and four charter schools. This training also did not provide support for local scoring and fidelity of scoring accuracy during training. This lack of training and monitoring of scorer accuracy of the WSS by kindergarten teachers may have introduced construct-irrelevant variance. Construct-irrelevant variance indicated some variance in the test-takers scores that could be attributable to extraneous factors that distort the meaning of the WSS scores and thereby decrease the potential validity of the proposed interpretation (AERA, APA, NCME, 2014).

Another potential limitation of this study was the establishment of the categorical cut scores of the FLKRS WSS. The authors of the *Standards for Educational and Psychological Testing* (2014) provide guidance on the rationale, procedures, and score interpretations related to cut scores. Standard 5.21 states that when proposed score interpretations involve one of more cut scores, the rationale and procedures used for establishing cut scores should be documented clearly. The WSS had a great deal of continuous data that were derived from 47 items. These continuous values were reduced to three discrete categorical variables: (1) *Not Yet*; (2) *In Process*; and (3) *Proficient*. It is important to note that in order to meet statute 1002.69(1), F.S., the Bureau of Just Read, Florida! condensed the three categories established by the authors of the WSS, (1) *Not Yet*; (2) *In Process*; and (3) *Proficient*, down to two discrete categorical variables: (0) *Not Ready* and (1) *Ready (In Process + Proficient)*. The process of translating continuous data elements into three categorical values inevitably caused the loss of information. Although the Bureau of Just Read, Florida! provided guidance that the original WSS cut score of *Not Yet* would equate to *Not Ready* and that the original cut scores of *In Process* and *Proficient* would

equate to *Ready* on the FLKRS WSS, psychometric explanations of how the new cut scores were established and validated have not been provided.

The original research questions were focused on the predictive relationships of determining overall kindergarten readiness on the WSS. The operational definition provided by the Bureau of Just Read, Florida! for the FLKRS WSS was *Ready* and *Not Ready*. However, when the data were analyzed for both cohorts of this study, the sample sizes for children earning cut scores of *Not Ready* were too small for reliable generalizations ($N = 13$ or 3.6% and $N = 19$ or 3.8%). In turn, the analysis of this study was changed to examine the predictive relationships of the levels of kindergarten readiness by using the original WSS cut scores of (1) *Not Yet*; (2) *In Process*; and (3) *Proficient*.

Research has shown that issues with the criterion within criterion-related validity studies is a common issue. Specifically, the identification of the criterion and how the said criterion should be measured. The FLKRS WSS may be losing information driven by policy and statute by the state of Florida by taking the continuous scores and placing these scores into categorical scores. With this loss of information and lack of psychometric support of the establishment of cut scores, the results of this study could be indicating that the independent variables of the VPK Assessment and Teaching Strategies GOLD® may be psychometrically working well, but the outcome measure itself is not working well. This is another reason that the original study was modified to use the original WSS cut scores to determine the level of kindergarten readiness. However, the study cannot discern if the outcome measure is working adequately as an operational definition of kindergarten readiness has not been provided by the State of Florida.

A final limitation to this study relates to Standard 7.12, which states that when test scores are used to make predictions about future behavior, the evidence supporting those predictions should be provided to the test user (AERA, APA, NCME, 2014). As stated previously, the Bureau of Just Read, Florida! has not provided evidence supporting that the cut scores from the FLKRS WSS validly represent kindergarten readiness for children across the state of Florida. Data have been shared related to percentages of children ready vs. not ready based on the FLKRS WSS but results from validation studies have not been shared with stakeholders. Additionally, during the period that this study was being conducted, the Bureau of Just Read, Florida! selected a new assessment measure to determine kindergarten readiness. Beginning with the 2017-2018 school year, each child's computer-adaptive score on Renaissance STAR Early Literacy® is used as a measure of kindergarten readiness. As with the FLKRS WSS, no support or psychometric rationale was provided from the state of Florida for the rationale of cut score of 500 or higher on the Renaissance STAR Early Literacy® to represent kindergarten readiness. However, reviewing the Renaissance STAR Early Literacy® technical assistance report on benchmarks, cut scores and growth rates, authors reported for grades K – 5, said scores are based on norms set in 2014-2015. For the fall screening window of kindergarten, the 40th percentile was a scale score of 496, while the 50th percentile was a scale score of 522. The Office of Early Learning for the state of Florida has not provided in writing the rationale of selecting a scale score for readiness of kindergarten that is below the 50th percentile of computer-adaptive assessment.

Implications for Future Research and Policy

The results of this study indicated that the VPK Assessment (standards-based) and Teaching Strategies GOLD® (curriculum-embedded) assessment measure subscales

independently provided moderate to strong predictive validity arguments for kindergarten readiness in the state of Florida. Additionally, the study found that when examining to determine if gender or ethnicity moderated the relationships, no evidence of differential predictive validity was evident for with independent variable. Although the specific predictive and differential validity relationships were found within this study, validity is an ongoing process that warrants a continuing body of evidence to continue to identify and support relationships. In order to continue to examine the validity of these assessment measures, it is recommended that an ongoing local study be conducted to include item-level data for both the independent variables (VPK Assessment and Teaching Strategies GOLD®) and dependent variables (WSS™). The inclusion of item-level data would allow researchers to complete statistical models and tests to examine further other aspects of validity such as internal structure, response processes, relation to other variables and consequences of testing. With the addition of the item-level data, future studies could build upon the reliability and validity data previously obtained, but also focus on (1) item difficulty; (2) item discrimination; and (3) differential item functioning. By adding item-level data, not only can relationships be determined, but a deeper level of statistical analysis can be provided to see if fairness within testing is at issue.

The results of this research study provide multiple implications for future research. A similar predictive validation study between the VPK Assessment and Teaching Strategies GOLD® subscales and kindergarten readiness as determined by the FLKRS STAR Early Literacy® is needed. The FLKRS STAR Early Literacy® is a computer-adaptive assessment measure with 27 items that produces multiple scores (e.g., raw score, scale score, percentile rank, stanine). In order to meet the criterion of state statute, 1002.69(1), F.S., the state determined a scale score of 500 or greater would determine that the individual child was ready for

kindergarten. The study should be a secondary analysis of cohorts of blended Head Start/VPK classrooms from medium to large school districts across the state of Florida. By broadening the sample of the secondary analysis to other school districts, the results of the study can help to add to the research base and add to the ability to generalize the validity results across multiple groups.

It is important to remember that within the state of Florida, any child enrolled in VPK must be assessed at least two times a year (Fall and Spring) with the VPK Assessment measure. The assessment authors state that the purpose of the VPK Assessment measure is to predict a level of kindergarten readiness. The specific measure that the Bureau of Just Read, Florida! has selected to determine kindergarten readiness should provide all policy makers and stakeholders with psychometric evidence of a determination of readiness. This psychometric evidence has been absent from Department of Education publications related to FLKRS. Florida policy makers may consider requiring all kindergarten students to complete the STAR Early Literacy® two times a year to determine growth throughout the kindergarten year. By establishing this policy, the state of Florida would have computer-adaptive scale scores on children at both the beginning of the year and end of the year. Psychometric analyses could then be conducted to determine actual cut scores that indicate a level of readiness entering kindergarten within the state of Florida. Adding another kindergarten assessment window might increase pedagogical and policy buy-in of kindergarten teachers who often feel that the purpose of completing the FLKRS is to determine the effectiveness of VPK programs. Kindergarten teachers in turn should be using the results of the FLKRS to determine domains that children need specific intervention and extension through differentiated instruction. Finally, by the state of Florida collecting longitudinal follow-up of children classified within the three readiness categories on the FLKRS

WSS™ of *Not Yet, In Process* and *Proficient*, additional information could be gleaned into how the children differ on multiple variables (e.g., academic, social, behavioral) over time.

Additionally, research studies are needed to determine if potential policy changes are needed within the state of Florida for mandatory assessments for early childhood programs if not predicting success in kindergarten. As previously stated, predictive studies indicate the strength of the relationship between test scores and criterion scores that are obtained at a later time. Policy makers within the state of Florida need to conduct additional predictive studies with the state-mandated VPK Assessment and FLKRS measures. It is important for the studies to determine how accurately the VPK Assessment test scores predict the criterion performance of kindergarten readiness on the FLKRS. Based on the results of the studies, policy makers need to determine if statute 1002.69(1), F.S., needs to be revised with FLKRS results only used to determine kindergarten readiness and not the effectiveness of VPK providers. If this statute cannot be revised to solely focus on kindergarten readiness, state-level early education decision-makers must examine deeply the predictive validity of the mandated direct assessment in VPK, the VPK Assessment. Currently the state of Florida states that the results from the VPK Assessment should be used to develop learning goals and targets for four-year-old children to be successful in kindergarten based on the KEA results.

Conclusion

This study examined the predictive and differential relationships between the VPK Assessment and Teaching Strategies GOLD® and kindergarten readiness as determined by the Florida Kindergarten Readiness Screener (FLKRS) Work Sampling System™ (WSS). In order to examine test-criterion relationships, predictive and differential predictive relationships were explored for gender and ethnicity (Hispanic vs. Non-Hispanic) subgroups. For both the VPK

Assessment and Teaching Strategies GOLD® independent subscales and assessment measure robust predictive relationships were found. Additionally, no robust differential predictive relationships were found. These results indicated that there was no evidence that the assessment measures were predicting differently by subgroups in determining kindergarten readiness.

With regard to the predictive relationships of the VPK Assessment and Teaching Strategies GOLD® with kindergarten readiness, when looking at each assessment measure subscale independently, all showed robust predictive relationships with kindergarten readiness with the exception of the Cognitive subscale of Teaching Strategies GOLD®. When looking at all of the subscales of each measure together, none of the subscales indicated robust relationships with kindergarten readiness.

Limitations of the study highlighted the lack of generalizability with a convenience sample of one school district. Additionally, the fact that the FLKRS WSS measure had a large amount of continuous data parsed to three categorical values and finally to two categorical values caused the measure to lose a great deal of information. This loss of information could impact the manner in which the outcome measure of kindergarten readiness was working and therefore impact how the independent measures (VPK Assessment and Teaching Strategies GOLD®) worked in the model as well.

Further, additional research studies should be conducted to build upon the results of this predictive validity study. As each year, at least 75% of Florida's four-year-old children participate in VPK, examinations of the predictive relationship of the state-mandated VPK Assessment across multiple subgroups should continue. The results of these studies should provide policy makers in Florida with vital data for decision-making on the continued use of the VPK Assessment subscales and appropriate assessment practices within early childhood

classrooms for individualization and differentiated instructional practice. Additionally, as the use of Teaching Strategies GOLD® curriculum-embedded, performance measure continues to be widely used in Head Start programs and other preschool programs focusing on more formative assessment data, guidelines on specific subscales to hone in for differentiated instructional practices for kindergarten readiness are recommended.

REFERENCES

- Ackerman, D. J., & Barnett, W. S. (2005). *Prepared for kindergarten: What does "readiness" mean?.* Washington, DC: National Institute for Early Education Research.
- Administration for Children and Families. (2016). *Head start performance standards (45 CFR Chapter XIII).* Washington, DC: United States Department of Health and Human Services.
- American Federation of Teachers. (2003). *Where we stand: Early childhood education.* Washington, DC.: American Federation of Teachers.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* American Educational Research Association.
- Aragon, S., & Rowland, J. (2015). Governors' Top Education Issues: 2015 State of the State Addresses. ECS Education Trends. *Education Commission of the States.*
- Bagnato, S. J., & Macy, M. (2010). Authentic assessments in action: A “REAL” solution. *NHSA Dialog, 13(1), 42-45.*
- Barnett, W. S., Carolan, M. E., Squires, J. H., Clarke Brown, K., & Horowitz, M. (2015). The state of prekindergarten 2014: State preschool yearbook. *New Brunswick, NJ: National Institute for Early Education Research.*
- Bush, G.W. (2005, February 2). *State of the union address.* Washington, DC. Retrieved on July 1, 2015 from http://www.washingtonpost.com/wpsrv/politics/transcripts/bushtext_020205.html
- Caffrey, E., Fuchs, D., & Fuchs, L. S. (2008). The predictive validity of dynamic assessment: A review. *The Journal of Special Education, 41(4), 254-270.*
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences, 2nd Edition.* Routledge
- Committee for Economic Development. (2002). *Preschool for all: Investing in a productive and just society.* New York.
- Connors-Tadros, L. (2014). Information and Resources on Developing State Policy on Kindergarten Entry Assessment (KEA). CEELo FastFacts. *Center on Enhancing Early Learning Outcomes.*

- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281.
- Dennis, L.R., Rueter, J. A., & Simpson, C.G. (2013). Authentic assessment: Establishing a clear foundation for instructional practices. *Preventing School Failure: Alternative Education for Children and Youth*, 57(4), 189-195.
- Duncan, S. E., & De Avila, E. A. (1985). *Language assessment scales*. San Rafael, CA: Linguametrics Group.
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test* (4th ed.). Bloomington, MN: Pearson
- Early Childhood Learning and Knowledge Center. (2015a). Head Start program facts fiscal year 2014. Retrieved October 1, 2015, from <https://eclkc.ohs.acf.hhs.gov/sites/default/files/pdf/head-start-fact-sheet-fy-2014.pdf>
- Early Childhood Learning and Knowledge Center. (2015b). Head Start timeline. Retrieved October 1, 2015, from <https://eclkc.ohs.acf.hhs.gov/hslc/hs/50th-anniversary/docs/hs-50th-timeline.pdf>
- Florida Department of Education. (2011). *Florida VPK assessments technical manual*. Tallahassee, FL: Lonigan, C.J.: Florida State University.
- Florida Department of Education. (2011b). *Florida early learning and developmental standards for four-year-olds (2011)*. Tallahassee, FL: Florida Department of Education.
- Florida Department of Education. (2015). *Florida kindergarten readiness screener administration manual*. Tallahassee, FL: Department of Education.
- Florida Office of Early Learning. (2011). *Office of inspector general annual report*. Tallahassee, FL: MacKinnon, R.
- Florida Office of Early Learning. (2013). *Florida voluntary prekindergarten (VPK) assessment teacher's manual*. Tallahassee, FL: Office of Early Learning.
- Florida Office of Early Learning. (2014). *Florida kindergarten readiness screener (FLKRS) and voluntary prekindergarten (VPK) education program accountability: an overview October 2014*. Tallahassee, FL: Office of Early Learning.
- Florida Office of Early Learning. (2016). *Florida voluntary prekindergarten (VPK) assessment: an overview May 2016*. Tallahassee, FL: Office of Early Learning. Retrieved July 11, 2016, from <http://www.floridaearlylearning.com/sites/www/Uploads/files/VPK/VPK%20Assessment%202014-15%20Overview%20May%202016%20060316.pdf>
- Freedman, D. A. (2006). On the “So-called Huber Sandwich Estimator” and “Robust Standard Errors”. *The American Statistician*, 60(4), 299-302.

- Gallant, D. J. (2009). Predictive validity evidence for an assessment program based on the Work Sampling System in mathematics and language and literacy. *Early Childhood Research Quarterly, 24*(2), 133-141.
- Goldstein, J., & Flake, J.K. (2016). Towards a framework for the validation of early childhood assessment systems. *Educational Assessment, Evaluation, and Accountability, 28*, 273-293.
- Goldstein, J., McCoach, D.B., & Yu, H. (2017). The predictive validity of kindergarten readiness judgments: Lessons from one state. *The Journal of Educational Research, 110*(1), 50-60.
- Good Start, Grow Smart (2002):
<http://georgewbushwhitehouse.archives.gov/infocus/earlychildhood/earlychildhood.html>
- Halle, T., Zaslow, M., Wessel, J., Moodie, S., & Darling-Churchill, K. (2011). Understanding and choosing assessments and developmental screeners for young children: Profiles of selected measures. Retrieved from Office of Planning, Research, and Evaluation, Administration for Children and Families, US Department of Health and Human Services.
- Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure?. *Child Development, 76*(5), 949-967.
- Hanover Research. (2013). *Kindergarten entry assessment assessments: Practices and policies*. Washington, D.C.: Hanover Research
- Heckman, J. J (2011). Effective Child Development Strategies. In E. Zigler, W. S. Gilliam, & W. S. Barnett (Eds.), *The Pre-K Debates: Current Controversies & Issues* (2-8). Brookes Publishing Company. PO Box 10624, Baltimore, MD 21285.
- Heroman, C., Burts, D. C., Berke, K., & Bickart, T. S. (2010). *Teaching Strategies GOLD® objectives for development & learning: Birth through kindergarten*. Washington, DC: Teaching Strategies.
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.
- IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.
- Kagan, S. L., & Scott-Little, C. (2004). Early learning standards: Changing the parlance and practice of early childhood education?. *Phi Delta Kappan, 85*(5), 388-396.
- Kennedy-Salchow, S. (2005). *An Analysis of Florida's Voluntary Pre-K Program*. National Center for the Study of Privatization in Education.
- Kim, D. H., Lambert, R. G., & Burts, D. C. (2013). Evidence of the validity of Teaching Strategies GOLD® assessment tool for English language learners and children with disabilities. *Early Education & Development, 24*(4), 574-595.

- Kolker, J., Osborne, D., & Schnurer, E. (2004). Early child care and education: The need for a national policy. *Washington, DC: Center for National Policy.*
- Lambert, R. G., Kim, D. H., & Burts, D. C. (2015). The measurement properties of the Teaching Strategies GOLD® assessment system. *Early Childhood Research Quarterly, 33*, 49-63.
- Lonigan, C.J. Florida VPK Assessment Measures, Technical Manual. State of Florida, Department of Education, 2011.
- Macy, M., & Bagnato, S.J. (2010). Keeping it “REAL” with authentic assessment. *NHSA Dialog, 13(1)*, 1-20.
- Mashburn, A. J., & Henry, G. T. (2004). Assessing school readiness: Validity and bias in preschool and kindergarten teachers' ratings. *Educational Measurement: Issues and Practice, 23(4)*, 16-30.
- Maxwell, K. L., & Clifford, R. M. (2004). School readiness assessment. *YC Young Children, 59(1)*, 42.
- MacCarthy, D. (1972). *Manual for the McCarthy scales of children's abilities.* Psychological Corporation.
- McDermott, P. A., Leigh, N. M., & Perry, M. A. (2002). Development and validation of the preschool learning behaviors scale. *Psychology in the Schools, 39(4)*, 353-365.
- Mead, S. (2004). Open the Preschool Door, Close the Preparation Gap. PPI Policy Report. *Progressive Policy Institute.*
- Meisels, S. J. (2011). Common Core standards pose dilemmas for early childhood. *Washington Post*, 423-442.
- Meisels, S. J., Bickel, D. D., Nicholson, J., Xue, Y., & Atkins-Burnett, S. (2001). Trusting teachers' judgments: A validity study of a curriculum-embedded performance assessment in kindergarten to grade 3. *American Educational Research Journal, 38(1)*, 73-95.
- Meisels, S. J., Dichtelmiller, M. L., Jablon, J. R., Dorfman, A. B., & Marsden, D. B. (2001). *The work sampling system* (Vol. 5). Pearson Education/Pearson Early Learning. Retrieved 2014.
- Meisels, S. J., Liaw, F. R., Dorfman, A., & Nelson, R. F. (1995). The Work Sampling System: Reliability and validity of a performance assessment for young children. *Early Childhood Research Quarterly, 10(3)*, 277-296.
- Merrell, K. M. (2003). Preschool and Kindergarten Behavior Scales (PKBS-2). Retrieved October 27, 2006.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.
- Myers, R. H., & Myers, R. H. (1990). *Classical and modern regression with applications* (Vol. 2). Belmont, CA: Duxbury Press.

- Miles, S. B., & Stipek, D. (2006). Contemporaneous and longitudinal associations between social behavior and literacy achievement in a sample of low-income elementary school children. *Child Development*, 77(1), 103-117.
- Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén.
- National Center on Quality Teaching and Learning. *Ongoing Child Assessment: Overview*. Retrieved July 30, 2015, from, http://eclkc.ohs.acf.hhs.gov/hslc/tta-system/teaching/docs/OngoingAssessment_TeacherTips.pdf
- National Governor's Association. (2005). *Building the foundation for bright futures: Final report of the Task Force on School Readiness*. Retrieved May 18, 2015, from <http://www.nga.org>.
- Neuman, S. B., & Roskos, K. (2005). The state of state pre-kindergarten standards. *Early Childhood Research Quarterly*, 20(2), 125-145.
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).
- Obama, B. H. (2014, January 28). *State of the union address*. Washington, DC. Retrieved on July 1, 2015 from <https://obamawhitehouse.archives.gov/the-press-office/2014/01/28/president-barack-obamas-state-union-address>
- Panter, J. E., & Bracken, B. A. (2009). Validity of the Bracken School Readiness Assessment for predicting first grade readiness. *Psychology in the Schools*, 46(5), 397-409.
- Phillips, D. A., & Shonkoff, J. P. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development*. National Academies Press.
- Pierson, A. (2018). Exploring state-by-state definitions of kindergarten readiness to support informed policy making. Retrieved from <https://ies.ed.gov/ncee/edLabs/regions/northwest/blog/kindergarten-readiness.asp>
- Rathvon, N. (2004). *Early reading assessment: A practitioner's handbook*. Guilford Press.
- Russo, J. M., Williford, A. P., Markowitz, A. J., Vitello, V. E., & Bassok, D. (2019). Examining the validity of a widely-used school readiness assessment: Implications for teachers and early childhood programs. *Early Childhood Research Quarterly*, 48, 14-25.
- Teaching Strategies. (2013). *Teaching strategies gold® assessment system concurrent validity: an independent study comparing external assessment measures to teaching strategies gold®*. Washington, DC: Teaching Strategies.
- Sarama, J., DiBiase, A. M., Clements, D. H., & Spitler, M. E. (2004). The professional development challenge in preschool mathematics. *Engaging young children in mathematics: Standards for early childhood mathematics education*, 415-446.

- Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly*, 22(2), 173-187.
- Snow, C., & Van Hemel, S. (2008). *Early childhood assessment: Why what and how?* Washington, DC: National Academic Press.
- Snow, C., & Oh, S. S. (2011). Assessment in early literacy research. *Handbook of early literacy research*, 3, 375-395.
- Snow, K. L. (2006). Measuring school readiness: Conceptual and practical considerations. *Early education and development*, 17(1), 7-41.
- Snow, K. L. (2011). *Developing kindergarten readiness and other large-scale assessments systems: Necessary considerations in the assessment of young children*. Washington, DC: National Association for the Education of Young Children.
- Stipek, D. (2006). No child left behind comes to preschool. *The Elementary School Journal*, 106(5), 455-466.
- United States Census. (2015). American fact finder population index for county in Florida [Data file]. Retrieved from https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_15_SPL_K200201&prodType=table
- United States Department of Education. (2011). *Race to the Top – Early Learning Challenge application for initial funding (GFDA No. 84.412)*. Washington, DC.: Author
- United States Department of Education. (2013). *Frequently asked questions for the EAG-KEA 2013 competition for the FY 2012 funds*. Washington, DC.: Department of Education.
- United States Department of Education and United States Department of Health and Human Services. (2014). *The Race to the Top – Early Learning Challenge Year Two Progress Report*. Washington, DC.: Author.
- Urban Child Institute. (2012). *Kindergarten readiness and the future of Shelby county*. Retrieved on August 2, 2015, from, <http://www.urbanchildinstitute.org/articles/research-to-policy/overviews/kindergarten-readiness-and-the-future-of-shelby-county>
- Urban Child Institute. (2015). *Investments in early childhood program benefit our entire community*. Retrieved on August 2, 2015, from, <http://www.urbanchildinstitute.org/articles/features/investments-in-early-childhood-benefit-our-entire-community>
- Wat, A., Bruner, C., Hanus, A., Scott-Little, C., & Schultz, T. (2012). Kindergarten entry assessment (KEA): Race to the Top–Early learning challenge fund information (E1). *Boston, MA: Build Initiative*.

- Watson, S.D. (2011). The Right Policy at the Right Time: The Pew Prekindergarten Campaign. In E. Zigler, W. S. Gilliam, & W. S. Barnett (Eds.), *The Pre-K Debates: Current Controversies & Issues* (pp. 9-21). Brookes Publishing Company. PO Box 10624, Baltimore, MD 21285.
- Weisenfeld, G. (2017). Assessment tools used in kindergarten entry assessments (CEELO FastFact). New Brunswick, NJ: Center on Enhancing Early Learning Outcomes.
- Westfall, P. (2014) Kurtosis as peakedness, 1905 – 2014. R.I.P. *American Statistician*, 68(3): 191-195.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2007). *Woodcock-Johnson® III normative update*. Rolling Meadows, IL: Riverside.
- Zigler, E., Gilliam, W. S., & Barnett, W. S. (2011). *The Pre-K Debates: Current Controversies and Issues*. Brookes Publishing Company. PO Box 10624, Baltimore, MD 21285.

Appendicies

Appendix A: History of Head Start Funding and Implementation

Head Start. In 2015, Head Start celebrated 50 years of providing comprehensive services to program eligible children and families leading the way in early childhood experiences and school readiness. The fundamental premise of Head Start programs was to not only support the educational growth of young children, but to also support the mental, health, social and emotional development of children birth through five years. A tenet of Head Start has been to prioritize services in a manner that was responsive to each child and family's ethnic, cultural and linguistic heritage (ECLKC, 2015a). Additionally, Head Start has always emphasized and built on the foundation of the role of the parent as not only their child's first teacher, but also the most important. Head Start has required programs to build instrumental relationships with families that support positive parent-child relationships, family well-being and connections to peers and the community (ECLKCa).

In 2015, the United States Department of Health and Human Services (HHS) awarded Head Start grants to nearly 2,000 public and private nonprofit and for-profit agencies to provide local services. Head Start agency grantees were required to propose and implement individualized services to meet the specific needs of their local community and Head Start Performance Standards. Head Start programs offered a variety of services models, depending on the needs of the local community. Programs were based in centers, schools or family childcare homes. Head Start preschool services were half-day (three hours) or full-day (6 hours) dependent on the agencies funding sources. Some, cities, states and federal programs offer funding to expand Head Start to include more children within their communities. Although the 1994 reauthorization of Head Start, Congress established the creation of the Early Head Start program providing services for expectant mothers, infants and toddlers, the emphasis on

preschool aged children has continued within the early childhood community. Over 80 percent of children served within the Head Start/Early Head Start grants were children three- to four-year old. (ECLKC, 2015a).

Head Start: 1964 - 1975. In President Lyndon B. Johnson's first State of the Union Address in January 1964, he began his journey of building a *Great Society* by declaring a *War on Poverty*. The president's ultimate goal of the War on Poverty was to eradicate the systems and causes of poverty by creating job opportunities, increasing productivity and enhancing the overall quality of life. As President Johnson was a former teacher within a one-room schoolhouse in rural Texas, the president believed reverently that education was the key to eliminating the cycle of poverty. With the help of Sargent Shriver, the president assembled a committee of academic and civil rights activists to create the Economic Opportunity Act of 1964 to include such programs as Job Corps, Urban/Rural Community Action, VISTA, and Project Head Start. In 1965, President Johnson and Sargent Shriver enlisted Dr. Robert Cooke, a pediatrician from Johns Hopkins University to organize a committee of other pediatricians, sociologists and psychologists, including Edward Zigler of Yale University's Child Development program. This steering committee became focused on determining how to give disadvantaged children a head start through comprehensive education, health, nutrition and social services, with a dramatic emphasis on parent involvement (ECLKC, 2015b). Throughout these discussions, the name of project varied from *Kiddie Corps* and *Baby Corps*, with the hope the terminology *corps* would elicit activism within the political and social climates of the era. However, the name Project Head Start was ultimately chosen by the academics on the committee who understood that the achievement gap and that middle class students often were performing ahead of their lower socio-economic peers (Kagan, 2000). A key tenet of the

program was to be culturally responsive to the communities served and that the communities have an investment in its success through the contribution of parent and community volunteers (ECLKC, 2015a).

Project Head Start was implemented in the summers of 1965 and 1966 served over 560,000 children and families across the nation through eight-week summer demonstration projects in child development centers through the United States that provided medical care, dental care and mental health services (Kagan, 2003). Based on the success of the summer demonstration projects, Project Head Start is deemed a success in child development by The Government Accountability Office in 1969. At this time, Congress authorizes the newly named Head Start to be funded as a primarily part-day, nine-month program. At the end of the same year, as the Office of Economic Opportunity's major function was to innovate and develop new approaches to poverty, President Johnson decided to move Head Start to an established department, the Department of Health, Education and Welfare. However, a tremendous debate ensued related to whether to place Head Start within the Office of Education or within the Children's Bureau. Based on the comprehensive and broad mandates of Head Start, it was determined that Head Start would be assigned to the newly created Office of Child Development within the Children's Bureau. In 1970, President Nixon named original Project Head Start Steering Committee member, Dr. Edward Zigler, Chief of the Office of Child Development. In 1974, Congress officially transferred the Office of Child Development from the Office of Economic Opportunity to the Office of Health and Human Services.

Head Start celebrated its tenth anniversary by announcing over five million children has been served in nearly ten thousand centers in every state and territory. Not only did Head Start leaders communicate the merits of the program, so too did the national political and educational

leaders acknowledge the continued tenet of the program in providing individualized services based on the ever changing needs of the communities being served. In 1972, Congress amended the *Economic Opportunity Act* to expand Head Start program opportunities for children with a diagnosed handicap. This legislation mandated that at least ten percentage of Head Start's national enrollment be comprised of children with a diagnosed handicap. In order to provide the highest quality of service to the children, Head Start collaborated with the Office of Education Bureau of Education for the Handicapped to train and offer technical assistance to Head Start teachers.

At the end of the decade, Dr. Zigler began the Head Start Project Development Continuity in an effort to provide seamless transitions for not only Head Start children, but also children with identified handicapping disorders. This project was created to develop sequential and continuous system of providing educational and comprehensive services to children transitioning from Head Start to a version of the *Head Start Performance Standards*. This document provided detailed requirements and guidelines for each comprehensive service area for serving preschool aged children and their families. Dr. Zigler's vision of establishing Head Start performance standards expanded into a federal standards initiative for public schools within the United States (ECLKC, 2015b).

Head Start: 1976 - 1985. By Head Start's 20th anniversary, Dr. Zigler was continuing to praise the merits and success of the program's impact to children and families, increasingly discussing small class sizes and the incorporation of specific comprehensive services such as medical, dental, nutrition and social services. However, the second decade of Head Start's implementation had as many challenges as it had successes under the guidance of President Carter. In 1978, Congress approved President Carter's budget that included the first expansion

opportunities for existing Head Start programs with the caveat that the services provided to enrolled children must be as comprehensive as those in effect on the date of the original grant's enactment. Furthermore, Congress also required that all Head Start programs, regardless of accepting additional funding must adhere to the requirements and guidance provided in the Performance Standards published in 1975.

In 1979, President Carter seemed to continue his support of Head Start by convincing Congress to not only reauthorize the program for the next five years, as well as an allowance for a 20 percent funding increase each year in the number of children and families to be served. Yet, in that same year, President Carter recommended to move Head Start from the Office of Health and Human Services to the United States Department of Education. Under the guidance of past members of the original steering committee, the Head Start community nationwide vehemently communicated their disagreement with this move again postulating that one of the major differences and success of the program is the medical, dental, nutritional and social comprehensive services that Head Start provides not only to the enrolled child, but the entire family. Based on this fundamental function of the program, leaders shared with the president that the program should remain under the tutelage of Dr. Zigler's Children's Bureau within the Office of Health and Human Services. After much debate and discussion, President Carter decided not to pursue the notion of moving Head Start under the educational umbrella of federal management and mandates.

In 1980, Head Start continued to be an advocate for providing individualized services for children based on the needs of the specific communities, as well as society as a whole. Nearly a decade after Head Start program leadership worked to include a minimum of ten percent of a program's funded enrollment be reserved for children with an identified handicap, the program

moved forward in broadening the definition of handicapped to children with disabilities (ECLKC, 2015b).

With the inauguration of President Ronald Reagan in 1981, Head Start was once again aligned with a president that supported early childhood programs conceptually and through federal funding. In what was first viewed by Head Start supporters as a repeat of the previous administration's stance on early education, President Reagan required Head Start to no longer provide the eight-week summer sessions that was the basis of the original Project Head Start. Upon further guidance, the president mandated the Children's Bureau to convert the summer programs to school-year programs in order to increase the level of exposure enrolled children and families would incur will enrolled in Head Start. To support his position, President Reagan persuaded Congress to authorize an increase of 125 million dollars in funding to Head Start. In turn, the deletion of the summer option and increase in funding allowed for nearly ten thousand additional Head Start allocations across the county, which raised funded enrollment of Head Start children to nearly 4000,000 (ECLKC, 2015b).

All of President Reagan's support of Head Start was not just from a fiscal level. In 1983, in spite of continued threats to decrease funding at the federal level, the president approved essential training funds to ensure that Head Start centers hired and retained well-qualified and staff members. However, in order to convince Congress to provide said training funding, the concession was the establishment and implementation for specific criteria for utilizing the training funds at the local level. In addition to implementing practical applications for staff training, Dr. Zigler, the President and Congress continued to value and recognize the importance of comprehensive services both in the centers, especially those areas that could be easily transposed within a family's home.

As President Reagan's first term in office came to an end, his fiscal support of Head Start continued not to falter. The president was able to convince Congress to pass the *Human Services Reauthorization Act of 1984*. This reauthorization was through the 1986 fiscal year and ensured that the *Head Start Performance Standards* continued to be adhered to, training funds for staff continued and a new emphasis that children could now be eligible for up to two years of Head Start services. This was a programmatic shift in program enrollment as providing a continued level of early intervention in addition to just the year prior to kindergarten entry. In addition to the Head Start components within the Human Services Reauthorization Act of 1984, the President was also successful in persuading Congress to approve the second program expansion funding for existing Head Start programs. For the first time since 1965, the Head Start federal budget tops one billion dollars with the program serving slightly more than nine million children since the program's inception by Sargent Shriver and President Lyndon B. Johnson.

Head Start: 1986 – 1995. In President Reagan's second term of office, his support of Head Start remained unflappable, however, he witnessed a new debate between the continued increases in number of funded Head Start enrollment slots versus allowing children to receive multiple years of the comprehensive services the program provided. Dr. Zigler and the Head Start Bureau continued to communicate the vision of President Johnson and reach as many different children as possible with the Head Start experience. The bureaus placed a discrete emphasis on serving more families who had not been exposed to Head Start's comprehensive services, as well as increasing the number of days of operations for some programs. The bureau stated that this practice would allow the program to serve more children across the nation without requiring an increase in federal funding (ECLKC, 2015b). Although individuals could not argue with an option that provided comprehensive services to more children over time,

Congress did not follow the recommendation of the Head Start Bureau's stance on program options and in 1986 passed the *Human Services Reauthorization Act of 1986*. Within this act, Head Start was reauthorized through the 1990 fiscal year without an increase of funding and allowing children to continue to be served within the program for multiple years. Although continuity of services and care is a vital component of Head Start, program leaders viewed this decision as a defeat for the Head Start population.

Although the *Human Services Reauthorization Act of 1986* did not allow for funding increases to provide more enrollment opportunities for families, President Reagan continued to support Head Start's mission and vision when working with Congress. Continuing to recognize the importance of training for Head Start staff, congressional hearings were held on the need for quality educational experiences for the nation's youngest learners in elementary school and including Head Start. Congress realized that in order for children to continue to grow academically and socially with the world's children, our nation's learners needed exposure to high-quality teachers and staff to potentially improve the lives of children and families. In 1988, Congress appropriated more than 50 million dollars for Head Start salary improvements. In 1989, Congress continued to communicate the need for quality initiatives within the Head Start program by appropriating another 50 million dollars to be set aside from the booming economy to provide quality initiatives for Head Start in the future.

Soon after in 1990, President George H.W. Bush proposed to Congress the *Head Start Expansion and Quality Improvement Act* that was quickly passed. The millions of dollars that Congress had previously set aside for Head Start was mandated to be utilized within this expansion act with a minimum of two percent of a program's budget to be dedicated to teacher and staff training guidelines. Additionally, for the first time since the program's inception in

1965, the Act also set minimum requirements for Head Start teachers. Each program was now required to have at least one teacher in a Head Start classroom that held a Child Development Associate (CDA) certification or appropriate early childhood degree or credential (ECLKC, 2015b).

Two years later, under the watchful eye of President George H.W. Bush, Congress passed the *Head Start Improvement Act*. The Act included a funding increase of over 600 million dollars which allowed another 180,000 children and families to participate in Head Start services. Additionally, the Act solidified the governments focus on quality within the program by establishing transportation regulations, site reviews of grantees at the end of the first year of operation, provided health services for younger siblings of eligible children and child development training for parents. Also, in an effort to continually revise Head Start program eligibility criterion to meet the changing needs of society and local communities, local Head Start programs were encouraged by the federal government to target enrollment for those families with preschool aged children who were experiencing homelessness (ECLKC, 2015b). The objectives to this initiative were to: (1) enhance access of homeless families to Head Start services; (2) provide services responsive to the special needs of homeless children and families; (3) identify effective methods of addressing the needs of homeless families; and (4) implement strategies for continued collaboration between Head Start programs and other community resources and agencies (ECLKCb).

As President Clinton became to be known as the ‘President of the Middle Class’, young children from families from lower socio-economic families also strongly benefited from the president’s initiatives. At the beginning of his first term, President Clinton commissioned the Office of Health and Human Services and the Department of Education to provide policy makers

with a vision for early childhood and K-12 educational systems as the nation began preparing for the 21st century. In 1993, an advisory committee on Head Start quality and expansion released a report that called for continued expansion of Head Start in the upcoming reauthorization process in 1994, as well as recommending for the creation of the Early Head Start program. In 1994, the president supported and Congress agreed with the recommendations and passed the landmark *Head Start Reauthorization Act of 1994*. The Act continued to expand Head Start by appropriating almost four billion dollars with services to be expanded to over 750,000 children. The law called for Head Start to continue to make significant program improvements, continue to increase staff qualifications and the development and provision for a Cost of Living Adjustment (COLA) for staff members. The final provision for Head Start required the Head Start Performance Standards to be revised therefore improve the level of monitoring of the federal and local programs. Finally, the most groundbreaking provision within the Act was the formulation of the Early Head Start program. This program established comprehensive services for expectant women, infants and toddlers (ECLKCb). As the nation celebrated the 30th anniversary of Head Start, the first 68 Early Head Start grants serving infants and toddlers were implemented across the nation, as was the beginning of the first Head Start longitudinal study of children and families across 17 sites (N=3,000).

Head Start: 1996 – 2005. At the end of President Clinton’s first term, he hosted a conference focused on early childhood development and learning. The conference highlighted research and updated scientific findings related to brain development that continued to support Head Start’s goal of reaching children in the earliest years to provide eligible children with a strong and healthy foundation of academic and social success (ECLKC, 2015b). The president and early childhood advocates and scholars praised the Head Start-Childcare Partnership

initiative that was spawned from the conference. The premise of the partnership was for early childhood programs to provide high quality, comprehensive, full-day services to meet the needs of children and their parents. Although this partnership was highly touted and supported by educators and politicians through the next decade, the program did not receive actual funding until 2015.

As the 1990s came to end, an era of level or reduced funding of federal programs seemed to be the norm. However, Congress passed the *Head Start Improvement Act of 1998*. Within the Act, Head Start was awarded a 411 million dollar; 374 million dollar; 311 million dollar and 608 million dollar increase in fiscal years 1997, 1998, 1999 and 2000, respectively. Head Start leaders and the federal and local level realized that with this increase in funding, Head Start operations would also begin to be influenced more by the K-12 educational system requirements. The Act began to revise the purpose of Head Start from a program that focused on the development of social competence to the promotion of school readiness. The Office of Health and Human Services not only revised the *Head Start Performance Standards* in order to include requirement and guidance for infants, toddlers and expectant mothers, but also focused on increased federal oversight at the local level and requested models of progress monitoring children's academic and social growth within the program. Finally, the law also highlighted the continued dialogue and program requirements for ensuring that high-quality staff was employed to impart social and academic knowledge to enrolled children. In each classroom, at least one teacher should have at minimum an Associate degree, rather than the minimum criteria of a high school diploma with a Child Development Associate credential (CDA).

As the White House transitioned from the Clinton to the George W. Bush administration, support for early childhood initiatives continued, though the sustained focus on cognitive

abilities over social competencies and comprehensive services also grew to the forefront of change. In 2001, First Lady Laura Bush, a former school librarian, hosted the White House Summit on Early Childhood Cognitive Development. Within this gathering of individuals, the First Lady's experts outlined what parents, grandparents, early childhood educators and childcare providers could systematically do to provide all children with rich experiences that primarily focused on early literacy skills. Data reported that at the beginning of the 21st century, more than 62 percent of young children received care through a variety of settings, including care by non-parental relatives, non-relatives, and center-based programs, including Head Start (White House, 2015a). Due to the fact that a majority of young children across the nation received child care outside of the home, federal and state governments provided more than 18 billion dollars to help families, in particular, children of families of lower socio-economic stature obtain non-parental care for children. Summit leaders reported that despite significant fiscal resources provided throughout the country, not all children were receiving high quality care. Three major reasons were highlighted: (1) most states had limited alignment between what children were doing before they enter school and what is expected of them once they enter the K-12 educational system; (2) early childhood programs were rarely monitored and evaluated based on how they prepared children to succeed in school; and (3) not enough information was available for early childhood teachers, parents, grandparents and childcare providers on methods of preparing children to be successful in school .

This summit gave birth to the Bush administration's proposal for the *Good Start, Grow Smart Initiative* in 2002. Within the initiative, President Bush outlined a critical step in educational reform with the need to prepare children to read and succeed in school with improved Head Start and early childhood development programs. The *Good Start, Grow Smart*

Initiative was designed to strongly align with the *No Child Behind Act* (Bush, 2001) that outlined similar academic expectations and monitoring systems for the K-12 public school population.

President Bush's administration highlighted three major areas to provide children with the skills they will need to enter K-12 educational system with an equal chance at achievement so that no child is left behind. The *Good Start, Grow Smart Initiative* key areas were: (1) strengthening Head Start; (2) partnering with states to improve early childhood education; and (3) providing information to teachers, caregivers and parents. The administration's plan to improve Head Start focused on requiring the Office of health and Human Services to create a Child Outcomes Framework to support programs plan their curricular activities and assess their children's progress and accomplishments of children (ECLKC b). The framework was comprised of eight early childhood general domains, 27 domains elements and a variety of examples of specific indicators of children's skills, abilities, knowledge and behaviors.

The Bush administration proposed stronger federal-state partnerships to ensure the delivery of quality early learning programs. The administration believed that efforts to improve early childhood experiences would not be successful without the involvement of states and local school districts. The underlying principle was that as states and districts were directly responsible for student learning and achievement in the K-12 schools, preparing children to learning before they enter school would be in their best interest (Bush, 2005). President Bush proposed that states take steps to help children prepare children before they enter school to be ready to learn with states coordinating with public school with local early childhood programs that serve the children they later will serve. The administration stated that this partnership needed to be accomplished through making available to early childhood programs information on what was expected of children once they reach school and the skills needed to meet K-12

state standards. This strengthened partnership asked states to develop quality criteria for early childhood education, including voluntary guidelines on emergent literacy and language skills and activities that were aligned with the state's K-12 education standards that were adaptable across various childcare settings (Bush).

Head Start: 2006 – 2015. In 2007, Head Start was once again brought to Congress for reauthorization. The 110th Congress passed and President Bush signed the *Improving Head Start for School Readiness Act of 2007* (H.R. 1429). This law formalized many items President Bush highlighted within the *Good Start, Grow Smart Initiative*. The new law made quality improvements in all comprehensive areas of Head Start including education staff qualifications, monitoring and collaboration with an increased level of accountability through continued federal monitoring and oversight. The law also continued to outline the desired outcome of promoting coordination among Head Start grantees and other state and local early childhood programs, including implemented blended funding with newly created state-funded universal preschool opportunities and state early childhood standards.

Appendix B: Work Sampling System P4, 5th Edition Scoring Tables

WSS Rating	Point Value
NY (Not Yet)	1
IP (In Process)	2
P (Proficient)	3

Florida Department of Education (2015)

Domain	Number of Indicators Rated	Not Yet	In Process	Proficient
Personal and Social Development	12	12-20	21-32	33-36
	11	11-18	19-29	30-33
	10	10-17	18-27	28-30
	9	9-15	16-24	25-27
Language and Literacy	10	10-17	18-27	28-30
	9	9-15	16-24	25-27
	8	8-13	14-21	22-24
Mathematical Thinking	12	12-20	21-32	33-36
	11	11-18	19-29	30-33
	10	10-17	18-27	28-30
	9	9-15	16-24	25-27
Scientific Thinking	4	4-6	7-10	11-12
	3	3-4	5-7	8-9
Physical Development, Health, and Safety	7	7-11	12-18	19-21
	6	6-10	11-16	17-18
	5	5-8	9-13	14-15
TOTAL	45	45-77	78-122	123-135
	44	44-76	77-120	121-132
	43	43-74	75-117	118-129
	42	42-73	74-115	116-126
	41	41-71	72-112	113-123
	40	40-69	70-109	110-120
	39	39-67	68-106	107-117
	38	38-66	67-104	105-114
	37	37-64	65-101	102-111
	36	36-62	63-98	99-108
	35	35-60	61-95	96-105
	34	34-59	60-93	94-102

Florida Department of Education (2015)

Appendix C: Sample of Mplus Code of Huber-White Sandwich Estimator

TITLE: Kindergarten Regression analyses

DATA:

File = E:\DECAMILLA_2014.txt;

Variable:

Names are

SITE, CLASS, PC, PA, MS, OL, SE, COG, LIT, MATH, WSS, GENDER, RACE,

ETHXPC,

ETHXPA,

ETHXMS,

ETHXOL,

ETHXSE,

ETHXCOG,

ETHXLIT,

ETHXMATH,

GENXPC,

GENXPA,

GENXMS,

GENXOL,

GENXSE,

GENXCOG,

GENXLIT,

GENXMATH;

usevariables are CLASS, PA, WSS, GENDER, RACE, GENXPA;

categorical are WSS;

missing are all (999);

cluster = CLASS;

Analysis:

type=complex;

Model:WSS ON PA, GENDER, RACE, GENXPA;

Output: sampstat modindices residual stand;

Appendix D: Ordinal Logistic Regression Model Summary Tables

Table A1.

Logistic Models Predicting Kindergarten Readiness with VPK Assessments—Cohort 1

	β	SE β	$\beta / SE \beta$	<i>p</i>
Model 1				
Gender (0 = Female, 1 = Male)	-0.35	0.10	-3.66	.00
Ethnicity (0 = Hispanic, 1 = Non-Hispanic)	0.11	0.11	1.05	.30
Model 2				
Gender	-0.30	0.10	-3.10	.00
Ethnicity	0.03	0.11	0.25	.80
Print Knowledge	0.01	0.02	0.31	.76
Phonological Awareness	0.01	0.02	0.73	.47
Mathematics	0.02	0.01	1.96	.05
Oral Language/Vocabulary	-0.01	0.02	-0.22	.83
Model 3a				
Gender	-0.64	0.27	-2.36	.18
Ethnicity	0.02	0.11	0.21	.84
Print Knowledge	0.03	0.03	0.94	.35
Phonological Awareness	0.01	0.02	0.46	.65
Mathematics	0.03	0.03	1.31	.19
Oral Language/Vocabulary	0.02	0.02	0.81	.42
Gender X Print Knowledge	0.04	0.03	1.33	.18
Model 3b				
Gender	-0.34	0.27	-1.29	.20

Table A1 Continued

Ethnicity	0.03	0.11	0.24	.81
Print Knowledge	0.05	0.03	1.88	.06
Phonological Awareness	0.01	0.02	0.37	.71
Mathematics	0.03	0.03	1.32	.19
Oral Language/Vocabulary	0.01	0.02	0.74	.46
Gender X Phonological Awareness	0.01	0.03	0.18	.86

Model 3c

Gender	-0.84	0.29	-2.92	.00
Ethnicity	0.02	0.11	0.18	.86
Print Knowledge	0.05	0.03	1.82	.07
Phonological Awareness	0.01	0.02	0.47	.64
Mathematics	0.01	0.03	0.22	.82
Oral Language/Vocabulary	0.02	0.02	0.83	.41
Gender X Mathematics	0.05	0.02	1.94	.05

Model 3d

Gender	-0.40	0.52	-0.77	.44
Ethnicity	0.03	0.11	0.24	.81
Print Knowledge	0.05	0.03	1.90	.06
Phonological Awareness	0.01	0.02	0.49	.62
Mathematics	0.03	0.03	1.32	.19
Oral Language/Vocabulary	0.01	0.03	0.42	.67
Gender X Oral Language/Vocabulary	0.01	0.03	0.21	.84

Table A1 Continued

Note: Due to larger sample size, statistical significance determined with a p -value $\leq .01$; Models 3a-3d build upon Model 2 by adding interactions of Gender and individual VPK Assessment subscales

Table A2.

Logistic Models Predicting Kindergarten Readiness with VPK Assessment-Cohort 1

	β	SE β	$\beta / SE \beta$	p
Model 4a				
Gender	-0.30	0.09	-3.20	.00
Ethnicity	-0.24	0.29	-0.84	.40
Print Knowledge	0.04	0.03	1.03	.30
Phonological Awareness	0.01	0.02	0.47	.64
Mathematics	0.03	0.03	1.29	.20
Oral Language/Vocabulary	0.01	0.02	0.77	.44
Ethnicity X Print Knowledge	0.03	0.03	1.02	.31
Model 4b				
Gender	-0.30	0.09	-3.18	.00
Ethnicity	-0.06	0.32	-0.19	.85
Print Knowledge	0.05	0.03	1.89	.06
Phonological Awareness	0.01	0.03	0.20	.84
Mathematics	0.03	0.03	1.26	.21
Oral Language/Vocabulary	0.01	0.02	0.76	.45
Ethnicity X Phonological Awareness	0.01	0.03	0.29	.77

Table A2 Continued

Model 4c

Gender	-0.30	0.09	-3.19	.00
Ethnicity	-0.26	0.32	-0.82	.42
Print Knowledge	0.05	0.03	1.88	.06
Phonological Awareness	0.01	0.02	0.46	.65
Mathematics	0.02	0.03	0.53	.59
Oral Language/Vocabulary	0.02	0.02	0.82	.41
Ethnicity X Mathematics	0.02	0.03	0.95	.34

Model 4d

Gender	-0.29	0.10	-3.09	.00
Ethnicity	0.34	0.56	0.61	.54
Print Knowledge	0.05	0.03	1.93	.05
Phonological Awareness	0.01	0.02	0.51	.61
Mathematics	0.03	0.03	1.95	.05
Oral Language/Vocabulary	-0.01	0.03	0.86	.39
Ethnicity X Oral Language/Vocabulary	-0.02	0.03	-0.59	.56

Note: Due to larger sample size, statistical significance determined with a p -value $\leq .01$; Models 4a-4d build upon Model 2 by adding interactions of Ethnicity and individual VPK Assessment subscales

Table A3.

Logistic Models Predicting Kindergarten Readiness with VPK Assessments-Cohort 2

	β	SE β	$\beta / SE \beta$	<i>p</i>
Model 1				
Gender (0 = Female, 1 = Male)	-0.19	0.09	-2.07	.04
Ethnicity (0 = Hispanic, 1 = Non-Hispanic)	0.07	0.10	0.64	.52
Model 2				
Gender	-0.14	0.10	-1.37	.17
Ethnicity	-0.00	0.11	- 0.01	.99
Print Knowledge	0.02	0.02	0.85	.40
Phonological Awareness	-0.03	0.02	-1.13	.26
Mathematics	0.06	0.02	3.09	.00
Oral Language/Vocabulary	0.04	0.02	1.99	.05
Model 3a				
Gender	-0.19	0.32	-0.57	.57
Ethnicity	0.01	0.11	0.05	.96
Print Knowledge	0.02	0.03	0.73	.47
Phonological Awareness	-0.03	0.02	-1.14	.26
Mathematics	0.06	0.02	2.98	.00
Oral Language/Vocabulary	0.05	0.02	2.04	.04
Gender X Print Knowledge	0.01	0.03	0.16	.87
Model 3b				
Gender	-0.20	0.32	-0.62	.54
Ethnicity	0.01	0.11	0.07	.95

Table A3 Continued

Print Knowledge	0.02	0.02	0.99	.32
Phonological Awareness	-0.03	0.03	-1.07	.29
Mathematics	0.06	0.02	2.99	.00
Oral Language/Vocabulary	0.05	0.02	2.04	.04
Gender X Phonological Awareness	0.01	0.03	0.20	.84

Model 3c

Gender	0.01	0.33	0.02	.99
Ethnicity	0.01	0.11	0.08	.94
Print Knowledge	0.02	0.02	1.02	.31
Phonological Awareness	-0.03	0.03	-1.17	.24
Mathematics	0.07	0.02	2.81	.00
Oral Language/Vocabulary	0.05	0.02	2.06	.04
Gender X Mathematics	-0.01	0.03	-0.44	.66

Model 3d

Gender	-0.02	0.62	-0.03	.97
Ethnicity	0.01	0.11	0.06	.96
Print Knowledge	0.02	0.02	1.00	.32
Phonological Awareness	-0.03	0.02	-1.14	.25
Mathematics	0.06	0.02	3.02	.00
Oral Language/Vocabulary	0.05	0.03	1.85	.06
Gender X Oral Language/Vocabulary	-0.01	0.03	-0.17	.87

Table A3 Continued

Note: Due to larger sample size, statistical significance determined with a p -value $\leq .01$; Models 3a-3d build upon Model 2 by adding interactions of Gender and individual VPK Assessment subscales

Table A4.

Logistic Models Predicting Kindergarten Readiness with VPK Assessments-Cohort 2

	β	SE β	$\beta / SE \beta$	p
Model 4a				
Gender	-0.13	0.11	-1.20	.23
Ethnicity	0.53	0.30	1.78	.08
Print Knowledge	0.06	0.03	2.07	.04
Phonological Awareness	-0.03	0.03	-1.09	.28
Mathematics	0.06	0.02	2.92	.00
Oral Language/Vocabulary	0.04	0.02	1.91	.06
Ethnicity X Print Knowledge	-0.06	0.03	-1.86	.06
Model 4b				
Gender	-0.14	0.11	-1.30	.20
Ethnicity	0.36	0.26	1.34	.18
Print Knowledge	0.03	0.02	1.10	.27
Phonological Awareness	-0.00	0.03	-0.11	.92
Mathematics	0.06	0.02	2.95	.00
Oral Language/Vocabulary	0.04	0.02	1.95	.05
Ethnicity X Phonological Awareness	-0.04	0.03	-1.32	.19

Table A4 Continued

Model 4c

Gender	-0.14	0.11	-1.26	.21
Ethnicity	0.69	0.26	2.59	.01
Print Knowledge	0.02	0.02	1.05	.30
Phonological Awareness	-0.03	0.03	-1.20	.23
Mathematics	0.10	0.03	3.71	.00
Oral Language/Vocabulary	0.04	0.02	1.87	.06
Ethnicity X Mathematics	-0.06	0.02	-2.61	.01

Model 4d

Gender	-0.14	0.10	-1.30	.19
Ethnicity	0.45	0.45	0.99	.32
Print Knowledge	0.02	0.02	1.03	.30
Phonological Awareness	-0.03	0.02	-1.19	.24
Mathematics	0.06	0.02	2.99	.00
Oral Language/Vocabulary	0.06	0.03	2.33	.02
Ethnicity X Oral Language/Vocabulary	-0.02	0.03	-0.95	.34

Note: Due to larger sample size, statistical significance determined with a p -value $\leq .01$; Models 4a-4d build upon Model 2 by adding interactions of Ethnicity and individual VPK Assessment subscales

Table A5.

*Logistic Models Predicting Kindergarten Readiness with Teaching Strategies GOLD®**- Cohort 1*

	β	SE β	$\beta / SE \beta$	<i>p</i>
Model 1				
Gender (0 = Female, 1 = Male)	-0.35	0.10	-3.66	.00
Ethnicity (0 = Hispanic, 1 = Non-Hispanic)	0.11	0.11	1.05	.30
Model 2				
Gender	-0.25	0.11	-2.27	.02
Ethnicity	-0.02	0.13	-0.15	.89
Social Emotional	0.01	0.02	0.31	.76
Cognitive	0.01	0.02	0.73	.47
Literacy	0.02	0.01	1.96	.05
Mathematics	-0.01	0.02	-0.22	.83
Model 3a				
Gender	-0.67	0.77	-0.86	.39
Ethnicity	-0.02	0.13	-0.16	.88
Social Emotional	-0.01	0.02	-0.52	.60
Cognitive	0.01	0.02	0.72	.47
Literacy	0.02	0.01	1.92	.06
Mathematics	-0.01	0.03	-0.21	.83
Gender X Social Emotional	0.01	0.01	0.54	.59
Model 3b				
Gender	-0.31	0.67	-0.46	.64

Table A5 Continued

Ethnicity	-0.02	0.13	-0.15	.88
Social Emotional	-0.00	0.02	-0.31	.76
Cognitive	0.01	0.02	0.73	.47
Literacy	0.02	0.01	1.94	.05
Mathematics	-0.01	0.03	-0.22	.83
Gender X Cognitive	0.00	0.01	0.09	.93

Model 3c

Gender	-0.95	0.44	-2.14	0.03
Ethnicity	-0.02	0.13	-0.17	0.86
Social Emotional	-0.00	0.02	-0.26	0.80
Cognitive	0.01	0.02	0.70	0.49
Literacy	0.02	0.01	1.51	0.13
Mathematics	-0.01	0.03	-0.29	0.78
Gender X Literacy	0.01	0.01	1.51	0.13

Model 3d

Gender	-0.83	0.55	-1.51	.13
Ethnicity	-0.02	0.13	-0.14	.89
Social Emotional	-0.00	0.02	-0.26	.80
Cognitive	0.01	0.02	0.70	.49
Literacy	0.03	0.01	-1.52	.13
Mathematics	-0.02	0.03	-0.54	.59
Gender X Mathematics	0.02	0.02	1.03	.30

Table A5 Continued

Note: Due to larger sample size, statistical significance determined with a p -value $\leq .01$; Models 3a-3d build upon Model 2 by adding interactions of Gender and individual VPK Assessment subscales

Table A6.

Logistic Models Predicting Kindergarten Readiness with Teaching Strategies GOLD® - Cohort 1

	β	SE β	$\beta / SE \beta$	p
Model 4a				
Gender	-0.23	0.11	-2.31	.02
Ethnicity	-1.01	0.82	-1.23	.22
Social Emotional	-0.02	0.02	-0.79	.43
Cognitive	0.01	0.02	0.76	.45
Literacy	0.02	0.01	1.89	.06
Mathematics	-0.00	0.03	-0.18	.86
Ethnicity X Social Emotional	0.02	0.01	1.28	.20
Model 4b				
Gender	-0.25	0.11	-2.28	.02
Ethnicity	-0.08	0.80	-0.10	.92
Social Emotional	-0.01	0.02	-0.31	.76
Cognitive	0.01	0.02	0.62	.54
Literacy	0.02	0.01	1.96	.05
Mathematics	-0.01	0.02	-0.22	.83
Ethnicity X Cognitive	0.00	0.01	0.10	.92

Table A6 Continued

Model 4c

Gender	-0.26	0.11	-2.30	.02
Ethnicity	-0.21	0.53	-0.40	.69
Social Emotional	-0.01	0.02	-0.32	.75
Cognitive	0.01	0.02	0.73	.46
Literacy	0.02	0.01	1.60	.11
Mathematics	-0.01	0.02	-0.23	.82
Ethnicity X Literacy	0.00	0.01	0.39	.69

Model 4d

Gender	-0.25	0.11	-2.28	.02
Ethnicity	-0.12	0.62	-0.19	.85
Social Emotional	-0.01	0.02	-0.31	.76
Cognitive	0.01	0.02	0.73	.47
Literacy	0.02	0.01	1.95	.05
Mathematics	-0.01	0.03	-0.28	.79
Ethnicity X Mathematics	0.00	0.02	0.17	.87

Note: Due to larger sample size, statistical significance determined with a p -value $\leq .01$; Models 4a-4d build upon Model 2 by adding interactions of Ethnicity and individual VPK Assessment subscales

Table A7.

Logistic Models Predicting Kindergarten Readiness with Teaching Strategies GOLD®
 - Cohort 2

	β	SE β	$\beta / SE \beta$	<i>p</i>
Model 1				
Gender (0 = Female, 1 = Male)	-0.20	0.09	-2.07	.04
Ethnicity (0 = Hispanic, 1 = Non-Hispanic)	0.07	0.10	0.64	.52
Model 2				
Gender	-0.22	0.10	-2.23	.03
Ethnicity	-0.02	0.13	0.19	.85
Social Emotional	0.01	0.02	0.31	.76
Cognitive	-0.02	0.02	-1.19	.23
Literacy	0.02	0.01	2.18	.03
Mathematics	0.01	0.02	0.57	.57
Model 3a				
Gender	-0.92	0.68	-1.34	.18
Ethnicity	-0.03	0.13	0.20	.84
Social Emotional	0.00	0.02	0.08	.94
Cognitive	-0.02	0.02	-1.17	.24
Literacy	0.02	0.01	2.19	.03
Mathematics	0.01	0.02	0.62	.54
Gender X Social Emotional	0.01	0.01	1.02	.31
Model 3b				
Gender	0.05	0.57	0.09	0.93

Table A7 Continued

Ethnicity	0.02	0.13	0.18	0.86
Social Emotional	0.01	0.01	0.76	0.45
Cognitive	-0.02	0.02	0.98	0.33
Literacy	0.02	0.01	2.15	0.03
Mathematics	0.01	0.02	0.56	0.58
Gender X Cognitive	-0.01	0.01	-0.50	0.62

Model 3c

Gender	-0.32	0.47	-0.67	0.50
Ethnicity	0.02	0.13	0.18	0.86
Social Emotional	0.01	0.01	0.73	0.47
Cognitive	-0.02	0.02	-1.19	0.24
Literacy	0.02	0.01	2.26	0.02
Mathematics	0.01	0.02	0.56	0.57
Gender X Literacy	0.00	0.01	0.21	0.83

Model 3d

Gender	-0.39	0.56	-0.70	.48
Ethnicity	0.02	0.13	0.19	.85
Social Emotional	0.01	0.01	0.73	.47
Cognitive	-0.02	0.02	-1.19	.24
Literacy	0.02	0.01	2.18	.03
Mathematics	0.01	0.02	0.38	.71
Gender X Mathematics	0.00	0.01	0.30	.76

Table A7 Continued

Note: Due to larger sample size, statistical significance determined with a p -value $\leq .01$; Models 3a-3d build upon Model 2 by adding interactions of Gender and individual VPK Assessment subscales

Table A8.

Logistic Models Predicting Kindergarten Readiness with Teaching Strategies GOLD®
- Cohort 2

	β	SE β	$\beta / SE \beta$	p
Model 4a				
Gender	-0.23	0.11	-2.18	.03
Ethnicity	2.04	0.88	2.32	.02
Social Emotional	0.03	0.02	1.94	.05
Cognitive	-0.02	0.02	-1.33	.18
Literacy	0.02	0.01	2.13	.03
Mathematics	0.01	0.02	0.60	.55
Ethnicity X Social Emotional	-0.03	0.02	-2.18	.03
Model 4b				
Gender	-0.23	0.11	-2.22	.03
Ethnicity	1.50	0.88	1.70	.09
Social Emotional	0.01	0.01	0.70	.49
Cognitive	-0.00	0.02	-0.03	.98
Literacy	0.02	0.01	2.13	.03
Mathematics	0.01	0.02	0.51	.61
Ethnicity X Cognitive	-0.03	0.02	-1.62	.11

Table A8 Continued

Model 4c

Gender	-0.22	0.11	-2.05	.04
Ethnicity	1.17	0.56	2.09	.04
Social Emotional	0.01	0.01	0.72	.47
Cognitive	- 0.02	0.02	-1.13	.26
Literacy	0.04	0.01	2.60	.01
Mathematics	0.01	0.02	0.52	.60
Ethnicity X Literacy	-0.02	0.01	-2.01	.05

Model 4d

Gender	-0.23	0.11	-2.10	.04
Ethnicity	1.48	0.78	1.90	.06
Social Emotional	0.01	0.01	0.78	.44
Cognitive	- 0.02	0.02	-1.16	.25
Literacy	0.03	0.01	2.21	.03
Mathematics	0.03	0.03	1.37	.17
Ethnicity X Mathematics	- 0.04	0.02	-1.87	.06

Note: Due to larger sample size, statistical significance determined with a *p*-value $\leq .01$; Models 4a-4d build upon Model 2 by adding interactions of Ethnicity and individual VPK Assessment subscales

Appendix E: Approval from School District Review Board

Approval from the School District Review Board was received on February 20, 2018 and verified by letter. This letter is not included within this document to protect the identity of the district.

Appendix F: Letter of Approval from University Institutional Review Board



RESEARCH INTEGRITY AND COMPLIANCE
Institutional Review Boards, FWA No. 00001669
12901 Bruce B. Downs Blvd., MDC035 • Tampa, FL 33612-4799
(813) 974-5638 • FAX (813) 974-7091

4/26/2018

Elizabeth DeCamilla
Teaching and Learning
2416 Fairview Road
Spring Hill, FL 34609

RE: Not Human Subjects Research Determination

IRB#: Pro00032661

Title: Predictive Validity of Standards-based and Curriculum-embedded Assessments for
Predicting Readiness at Kindergarten Entry

Dear E. DeCamilla:

The Institutional Review Board (IRB) has reviewed your application. Study activities involve analysis of de-identified data from a public source. As such, the activities do not meet the definition of research under USF IRB policy, and USF IRB approval and oversight are therefore not required.

While not requiring USF IRB approval and oversight, your study activities should be conducted in a manner that is consistent with the ethical principles of your profession. If the scope of your project changes in the future, please contact the IRB for further guidance.

If you will be obtaining consent to conduct your study activities, please remove any references to "research" and do not include the assigned Protocol Number or USF IRB contact information.

If your study activities involve collection or use of health information, please note that there may be requirements under the HIPAA Privacy Rule that apply. For further information, please contact a HIPAA Program administrator at (813) 974-5638.

Sincerely,

A handwritten signature in black ink that reads "John A. Schinka, Ph.D." in a cursive style.

John Schinka, Ph.D., Chairperson
USF Institutional Review Board