Psychometric Properties of the Florida VPK Assessments in Use State-Wide: Data from the 2011-2012 and 2012-2013 VPK Years

Prepared for the Florida Office of Early Learning by
The Florida Center for Reading Research
Florida State University

June 26, 2014

Psychometric Properties of the Florida VPK Assessments in Use State-Wide: Data from the 2011-2012 and 2012-2013 VPK Years

Executive Summary

The analyses described in this report were intended to address five questions concerning the psychometric functioning and accurate use of the VPK Assessments when used state wide by VPK providers.

Specifically, the analyses conducted were designed to address the following questions:

- Is the VPK Assessment functioning similarly to how it functioned in the field trial (in terms of scores, growth, and reliability) when used state wide?
- Do scores on the VPK Assessment predict scores on the kindergarten readiness measures similar to results of field trial using the 2012-2013 Florida Kindergarten Readiness Screener (FLKRS) results?
- Do some VPK providers do much better or much worse than expected based on the average patterns of results in the state-wide data?
- How do scores vary in terms of initial status and growth across the VPK year for different subgroups of children?
- Are there differences in patterns of scores overall or for subgroups between the year in which use of the VPK Assessment was voluntary and the year in which use of the VPK Assessment was mandatory?

Results of concurrent and predictive correlational analyses indicate that the VPK Assessment functions similarly to how it functioned in the field trial. Cross-time correlations between subtests of the VPK Assessment (i.e., AP1 - AP2, AP1 - AP3, AP2 - AP3) were similar in magnitude to those obtained in the original field trial, indicating that the specific assessments at the different assessment periods are measuring reliable variance in children's abilities and that differences between children's relative rank in the overall distribution of scores remains somewhat stable from assessment period to assessment period. Similarly, correlations between scores on the subtests of the VPK Assessment at different assessment periods and scores from the Florida Kindergarten Readiness Screener (FLKRS) and Florida Assessment for Insruction in Reading-Kindergarten (FAIR-K), which are administered at the beginning of children's kindergarten year, were similar between the original field trial and the data from the year that the VPK Assessment was used state wide.

Analyses of growth were conducted to determine the overall pattern of changes in children's skills across the VPK year and to determine how different child characteristics influenced rates of growth in oral language, print knowledge, phonological awareness, and math skills during the

VPK year. Across both years and on all four skill domains, there was significant growth in scores on the VPK Assessment during the VPK year. The skills children had at the beginning of the VPK year and the rate of growth during the VPK year was significantly affected by each of the child characteristics examined. In general, girls and older children started the VPK year with higher skills and experienced more growth in skills across the VPK year than did boys and younger children. However, this pattern was qualified by the finding that children with higher initial scores demonstrated slower growth overall--likely because they had already been exposed to activities and experiences that fostered the development of age-appropriate skills in oral language, early literacy, and early math.

Data from the 2011-2012 VPK year included more variables indexing child characteristics than did the data from the 2012-2013 VPK year, because those data were linked to information associated with children's scores on the FLKRS. For data from the 2011-2012 VPK year, most child characteristics were predictive of children's initial skills. Children classified as eligible for Free and Reduce-Price Lunch (FRPL), Limited English Proficiency (LEP), or Exceptional Student Services (ESE) started the VPK year with lower skills in all four skill domains than did children without these designations. Additionally, children who were non-white or Latino started the VPK year with lower skills in all four skill domains than did children classified as white. Rates of growth in all four skill areas were most affected by LEP and ESE status. Children designated as LEP or ESE experienced less growth during the VPK year than did children without these designations.

In addition to examining the psychometric characteristics of the VPK Assessment when used state wide and patterns of growth, analyses were conducted to determine the degree to which providers were administering the assessments in a way that yielded accurate scores. Results of several analyses conducted addressed this question. First, there was no evidence that a substantially higher percentage of children were showing more or less growth than would be expected based on the overall pattern of growth for children. For each subtest of the VPK Assessment and across both years, close to the expected 16% of children had scores indicating higher or lower than expected growth. Moreover, children who demonstrated this higher or lower rate of growth were not concentrated in particular VPK providers--as would be expected if an overall pattern of inaccurate administration or reporting of the assessment were occurring. There were fewer providers in the year in which use of the VPK Assessment was mandatory that reported scores indicating that 50% or more of their children made lower than expected progress, when compared to data reported by VPK providers the prior year when use of the VPK Assessment was voluntary. Whether this change was the result of differences in administration procedures for the assessment or differences in instruction cannot be determined.

Comparisons of initial status and growth of the skills measured by the VPK Assessment across the 2011-2012 and 2012-2013 VPK years did not reveal a pattern of differences that would suggest that the change from voluntary use to mandatory use altered the way the measure was administered. That is, levels of initial skills and amount of growth in skills were comparable across years. Direct comparisons of results from VPK providers with data from both years revealed very similar profiles, suggesting a high degree of stability in initial scores and growth from year to year. Moreover, analyses of absolute differences in scores for VPK Assessment subtests at each assessment period and difference in absolute score changes from AP1 to AP3

across the 2011-2012 VPK year and the 2012-2013 VPK year, revealed small year-to-year differences. Although these differences were statistically significant, they represent less than 1/10 of a standard deviation difference and had no consistent pattern in terms of the direction of the difference.

Overall, the results of the analyses conducted indicate that the strong psychometric properties in terms of reliability and validity that the VPK Assessment demonstrated in development work and the field trial continue when the assessment is used on a broad scale. Additionally, the analyses provided little evidence that the validity of the VPK assessment varied from the field trial to state-wide use or was affected by the change from voluntary use to mandatory use.

Psychometric Properties of the Florida VPK Assessments in Use State-Wide: Data from the 2011-2012 and 2012-2013 VPK Years

The Florida Voluntary Pre-Kindergarten (VPK) Assessment is a tool developed to be used by teachers in Florida's Voluntary Preschool Program. With this assessment teachers can measure children's knowledge of early academic skills, plan instructional activities, and monitor progress three times across the preschool year. The measure assesses oral language, print knowledge, phonological awareness, and mathematics. The measure was designed to be administered individually and directly by VPK teachers. The Florida VPK assessment was developed by the Florida Center for Reading Research, and the results of development and field studies provided a high level of support for the psychometric quality of the measure. Initial validation of the VPK assessment was reported in the VPK Assessment Measure Technical Manual.

Since the development of the Florida VPK Assessment, the measure became readily available to VPK providers across the state of Florida. During the 2011-2012 school year teachers were able to access testing materials, assess children at the beginning (AP1), middle (AP2) and end (AP3) of the preschool year, and report individual children's data to the state on a voluntary basis. However, during the 2012-2013 VPK year, the use of the Florida VPK assessment became mandatory for VPK providers as part of accountability testing. VPK providers were required to give the Florida VPK assessment and report scores for each child at AP1 and again at AP3.

Although the strong psychometrics of the VPK Assessment was established initially, it is important to examine the psychometric properties of the measure after it was taken to full scale and again after it became mandatory and part of accountability testing for schools. In the analyses described in this report, we examined how changes in the way the assessment was used may have impacted the psychometric properties and the accuracy of scores. To determine how the context of administration of the measure may have affected how it was administered and scored, we conducted a series of analyses to examine the possibility that teachers may have been less accurate in their testing and/or reporting of VPK assessment scores after it became linked with their performance. We examined if these differences were greater for specific VPK providers indicating teachers in these schools may not be utilizing the VPK assessment properly. This information could be useful in identifying preschools that may benefit from further professional development in administration of the VPK assessment.

The series of analyses conducted with data from the 2011-2012 and 2012-2013 VPK years was directed at addressing the following specific questions:

- Is the VPK Assessment functioning similarly to how it functioned in the field trial (in terms of scores, growth, and reliability) when used state wide?
- Do scores on the VPK Assessment predict scores on the kindergarten readiness measures similar to results of field trial using the 2012-2013 Florida Kindergarten Readiness Screener (FLKRS) results?
- Do some VPK providers do much better or much worse than expected based on the average patterns of results in the state-wide data?
- How do scores vary in terms of initial status and growth across the VPK year for different subgroups of children?

- Are there differences in patterns of scores overall or for subgroups between the year in which use of the VPK Assessment was voluntary and the year in which use of the VPK Assessment was mandatory?

Examining Psychometric Functioning of VPK Assessment when Used State-Wide

Data from the field trial of the VPK Assessments provided information concerning how the measure should function when used by teachers (i.e., scores at different assessment intervals, relations of scores between measures within time, relations of scores within measures across time, relation of scores to student characteristics). Using data from the 2011 - 2012 and 2012 - 2013 VPK year, we examined if these same relations are obtained when the VPK Assessments are used on a much broader scale than the field trial.

Cross-time Consistency

We first examined the cross-time correlations between subtests of the VPK assessment measures across the 3 assessment periods (note: correlations between scores on each subtest within assessment period are shown in Tables A1 and A2). Higher correlations across time support the reliability of each subtest of the VPK assessment, suggesting that each subtest is measuring a similar underlying trait across the assessment periods it was administered. It was expected that high correlations would be found between assessment periods indicating a consistency in children's scores across time. In other words, children scoring high on a subtest at one time point were expected to score high on a subtest at another time point as well. Table 1 shows the cross-time correlations from the field trail that were originally reported in the VPK assessment technical manual, Table 2 shows the cross-time correlations for each subtest using data from the 2011-2012 VPK year, and Table 3 shows the cross-time correlations for each subtest using data from the 2012-2013 VPK year.

Table 1. ***Correlations between VPK Subtest Scores at Different Assessment Periods as reported in the VPK assessment technical manual***

	Test Interval		
VPK Assessment Subtest	AP1 to AP2	AP1 to AP3	AP2 to AP3
Print Knowledge	.70	.59	.81
Phonological Awareness	.64	.56	.73
Oral Language	--	--	.73
Math	.76	.69	.82

Notes. $N = 957$. AP = Assessment Period. Oral language not administered at AP1. All correlations are significant at $p < .001$.

Table 2. ***Correlations between VPK Subtest Scores at Different Assessment Periods Using Data from the 2011-2012 School Year***

	Test Interval		
VPK Assessment Subtest	AP1 to AP2	AP1 to AP3	AP2 to AP3
Print Knowledge	.65	.48	.74
Phonological Awareness	.60	.48	.67
Oral Language	.72	.60	.69
Math	.71	.59	.76

Notes. AP= Assessment Period; From AP1 to AP2 $N = 33{,}996$; From AP1 to AP3 $N = 32{,}554$; From AP2 to AP3 $N = 33{,}429$. All correlations are significant at $p < .001$.

Table 3. ***Correlations between VPK Subtest Scores at Different Assessment Periods Using Data from the 2012-2013 School Year***

	Test Interval		
VPK Assessment Subtest	AP1 to AP2	AP1 to AP3	AP2 to AP3
Print Knowledge	.65	.50	.76
Phonological Awareness	.58	.46	.65
Oral Language	.72	.63	.71
Math	.71	.60	.76

Notes. AP= Assessment Period. From AP1 to AP2 $N = 110{,}336$; From AP1 to AP3 $N = 125{,}351$; From AP2 to AP3 $N = 107{,}650$. All correlations are significant at $p < .001$.

Overall, cross-time correlations were very similar across school years indicating that the VPK assessment maintained reliability when taken to full scale (in the 2011-2012 VPK year) and again after it became mandatory and part of accountability testing for schools (in the 2012-2013 VPK year).

Compared to the cross-time correlations reported in the VPK assessment technical manual, correlations from both the 2011-2012 and 2012-2013 VPK years were slightly lower, indicating slightly less reliability in these scores. Typically in the rigorous context of a research study more reliability is obtained on measures than when measures are used in more "real-world" settings. Teachers in the 2011-2012 and 2012-2013 VPK years were no longer supported by trained research staff and may have had less direct supervision and less ability to receive one-on-one instruction in administering the VPK assessment.

Also of note is that across subtests, correlations between AP1 and AP3 were lower than correlations between assessment periods that occurred in direct succession (AP1 & AP2 and AP2 & AP3). This pattern occurred in the data reported from the technical manual, from the 2011-2012 VPK year and from the 2012-2013 VPK year. This result is likely due to the fact that children show greater gains in academic skills across the entire school year than from the 3 to 4 months between successive assessment periods. It was expected that children would make greater gains in skills from AP1 to AP3 than from AP1 to AP2 (or from AP2 to AP3), and their scores would, therefore, be less highly correlated between these non-adjacent assessment points.

Validity

To examine the validity of the VPK assessment--how well it measured what was intended to measure--we examined the relations between scores on the VPK Assessments in the 2011-2012 school year and scores obtained on the kindergarten readiness measures (FLKRS) assessed in the fall of children's kindergarten year. It was expected that, similar to the findings reported in the VPK assessment technical manual, VPK assessment scores would be significantly correlated with the kindergarten readiness indices given at the beginning of the 2012-2013 kindergarten year. FLKRS measures include the Florida Assessment of Reading for Instruction (FAIR-K) and Early Childhood Observation System (ECHOS). Table 4 shows the correlations between VPK assessment and FLKRS scores reported in the VPK assessment technical manual, and Table 5 shows the same correlations from the 2011-2012 VPK year (FLKRS scores were not available for children in the 2012-2013 VPK assessment data).

As can be seen in Tables 4 and 5 (following pages), correlations between VPK assessment scores and kindergarten readiness scores were similar across the two data sets. Across time periods the VPK assessment was significantly and moderately correlated with both the FAIR-K and ECHOS scores. Generally, correlations were stronger for later assessment periods than for earlier ones. For example, nearly all the VPK assessment subscale scores and kindergarten screener scores correlated more highly at AP3 than at AP1. This was expected because the time that elapsed between the AP3, given towards the end of the preschool year, and the kindergarten readiness measures, given at the beginning of the kindergarten year, was shorter than the time elapsed between when the other VPK assessment periods took place and the beginning of the kindergarten year. These significant correlations indicate the VPK assessment is measuring a similar and overlapping skill as the kindergarten readiness measures. The results also indicate that the VPK measures provided valid measurement of pre-academic skills once it was used at scale.

Table 4. Convergent and discriminant correlations between subtests of the VPK assessment and FLKRS scores as reported in the VPK technical manual

	FLKRS Scores					
	ECHOS	Scores from FAIR-K				
VPK Assessment Subtest	Total Score	PRS	Letter Name	PA	LC	Vocabulary
Assessment Period 1						
Print Knowledge	.29	.43	.39	.41	.22	.31
Phonological Awareness	.20	.31	.19	.37	.29	.40
Math	.32	--	--	--	--	--
Assessment Period 2						
Print Knowledge	.26	.53	.56	.39	.15	.23
Phonological Awareness	.22	.39	.24	.44	.30	.37
Oral Language	.28	.34	.23	.36	.37	.50
Math	.31	--	--	--	--	--
Assessment Period 3						
Print Knowledge	.32	.55	.63	.35	.13	.20
Phonological Awareness	.29	.45	.28	.48	.31	.38
Oral Language	.23	.34	.21	.37	.49	.34
Math	.22	--	--	--	--	--

Notes. N = 1,008 at AP1; N = 914 at AP2; N = 898 at AP3. FLKRS = Florida Kindergarten Readiness Screener; ECHOS= Early Childhood Observation System; FAIR-K= Florida Assessment of Reading for Instruction; PRS= Probability of Reading Success; PA= Phonological Awareness; LC= Listening Comprehension. All correlations are significant at $p < .001$.

Table 5. Convergent and Discriminant correlations between subtests of the VPK Assessment and Kindergarten Readiness Subtest Scores using data from the 2011-2012 VPK year and the 2012-2013 Kindergarten year.

	FLKRS Scores					
	ECHOS	Scores from FAIR-K				
VPK Assessment Subtest	Total Score	PRS	Letter Name	PA	LC	Vocabulary
Assessment Period 1						
Print Knowledge	.22	.38	.33	.35	.23	.33
Phonological Awareness	.23	.32	.21	.36	.31	.39
Oral Language	.30	.39	.27	.39	.42	.54
Math	.27	.41	.33	.39	.30*	.38*
Assessment Period 2						
Print Knowledge	.27	.51	.50	.40	.24	.33
Phonological Awareness	.25	.41	.27	.44	.32	.41
Oral Language	.27	.39	.28	.37	.38	.50
Math	.29	.47	.40*	.42*	.28*	.35*
Assessment Period 3						
Print Knowledge	.24	.51	.56	.33	.19	.25
Phonological Awareness	.25	.44	.31	.44	.30	.39
Oral Language	.25	.37	.28	.34	.34	.44
Math	.29	.49	.45*	.40*	.26*	.33*

Notes. N = 26,188 at AP1; N = 24,420 at AP2; N = 23,892 at AP3. FLKRS = Florida Kindergarten Readiness Screener; ECHOS= Early Childhood Observation System; FAIR-K= Florida Assessment of Reading for Instruction; PRS= Probability of Reading Success; PA= Phonological Awareness; LC= Listening Comprehension. All correlations are significant at $p <$ .001.

To further examine the validity of the VPK assessment during the 2011-2012 school year, the correlations between VPK assessment subtests and kindergarten readiness measure subtests assessing similar aspects of academic development (convergent correlations) were compared to correlations between VPK assessment subtests and kindergarten readiness measure subtests assessing different aspects of academic development (discriminant correlations). For example, it was expected that the FAIR-K Letter Naming subtest would correlate more highly with the VPK Print Knowledge subtest than with the VPK Math subtest. These types of analyses provide evidence that specific subtests of the VPK assessment were indeed measuring what they purport to measure. They suggest scores on a specific subtest more closely relate to other measures assessing similar skills than measure assessing more distantly related skills. Consequently, we examined if and when convergent correlations were statistically higher than discriminant correlations at each time point.

At AP1, the correlation between VPK Oral Language and FAIR-K Listening Comprehension scores was significantly higher than the correlation between VPK Math and FAIR-K Listening Comprehension scores ($Z = 24.23$, $p < .001$). As seen in Table 5, the convergent and discriminant correlations relating to the FAIR-K Letter Name subtest and the FAIR-K Phonological Awareness subtest were not significantly different.

At AP2, all convergent correlations were significantly higher than discriminant correlations among VPK assessment and FAIR-K subtest scores examined. The correlation between VPK Print Knowledge and FAIR-K Letter Name scores were significantly higher than the correlation between VPK Math and FAIR-K Letter Name Scores ($Z = 22.35$, $p <.001$). The correlation between VPK Phonological Awareness and FAIR-K Phonological Awareness scores was significantly higher than the correlation between VPK Math and FAIR-K Phonological Awareness scores ($Z = 4.06$, $p < .001$). The correlation between VPK Oral Language and Fair-K Listening Comprehension scores was significantly higher than the correlation between the VPK Math and the Fair-K Listening Comprehension scores ($Z = 18.10$, $p < .001$). The correlation between VPK Oral Language and FAIR-K Vocabulary scores was significantly higher than the correlation between the VPK Math and FAIR-K Vocabulary scores ($Z = 28.66$, $p < .001$).

Similar to AP2, all AP3 convergent correlations were significantly higher than AP3 discriminant correlations among VPK and FAIR-K subtest scores. The correlation between VPK Print Knowledge and FAIR-K Letter Name scores was significantly higher than the correlation between VPK Math and FAIR-K Letter Name Scores ($Z = 23.37$, $p < .001$). The correlation between VPK Phonological Awareness and FAIR-K Phonological Awareness scores was significantly higher than the correlation between VPK Math and FAIR-K Phonological Awareness scores ($Z = 8.3$, $p >.001$). The correlation between VPK Oral Language and FAIR-K Listening Comprehension Scores was significantly higher than the correlation between VPK Math and FAIR-K Listening Comprehension Scores ($Z = 15.63$ $p <.001$). The correlation between VPK Oral Language and FAIR-K Vocabulary scores was significantly higher than the correlation between VPK Math and FAIR-K Vocabulary Scores ($Z = 21.24$, $p < .001$).

Overall most of the convergent correlations from AP2 and AP3 were significantly higher than the discriminant correlations in the same assessment periods, indicating that the VPK subtests are valid and measure what they were intended to measure when taken to scale.

Table 6. ***Summary of Multiple Regressions for VPK Assessment Measures Predicting Probability of Reading Success from FAIR-K as Reported in the VPK Technical Manual***

		Unique Variance Accounted for in Model				
Predictor Variables	Overall R^2	Print Knowledge	Phonological Awareness	Oral Language	Shared Variance	*N*
AP1	.21	.12	.02	--	.07	1088
AP2	.33	.16	.02	.01	.14	914
AP3	.38	.17	.04	.001[a]	.17	898

Notes. R^2 = Variance accounted for by model. AP = Assessment Period. Unless otherwise marked, all total and unique variance components were significant at $p < .001$; [a]$p = .33$;

Multiple regression analyses were used to examine the contribution of the VPK assessment measures to FAIR-K probability of reading success scores (PRS) and the unique contributions of each subtest of the VPK assessment to FAIR-K PRS scores. It was expected that VPK assessment scores would significantly predict how children scored on the overall measure of reading readiness, the FAIR-K PRS. The contributions of each subtest, at each assessment period, were compared to the regression coefficients from the same analyses reported in the VPK technical manual. Table 6 shows the multiple regression results from the VPK technical manual and Table 7 shows results from the same multiple-regression analyses completed on data from the 2011-2012 VPK year.

The overall variance accounted for, at each assessment period, was very similar in the VPK technical manual and in the 2011-2012 VPK year, indicating that when taken to scale, the VPK scores significantly contributed to prediction of overall kindergarten reading readiness. This result provides more support for the validity of the VPK assessment when taken to scale and indicates the VPK assessment measures a similar underlying ability as the FAIR-K.

For the 2011-2012 VPK year, the AP1 VPK assessments accounted for 20% of the variance in FAIR-K PRS scores; the AP2 VPK assessments accounted for 30% of the variance in FAIR-K PRS scores; and the AP3 VPK assessment accounted for 31% of the variance in FAIR-K PRS scores. Individual regression coefficients for each VPK subtest, at each assessment period, were similar across the two VPK years. In the 2011-2012 VPK year, all subtests significantly contributed to FAIR-K PRS scores. Although the shared variance at each assessment period was slightly higher in the 2011-2012 year than in the technical manual, the regression coefficients tended to be lower across all assessment periods. This result suggests that although, overall, the VPK assessment provides similar predictive utility when taken to scale, the individual subtest scores individually provide slightly less predictive information.

Table 7. ***Summary of Multiple Regressions for VPK Assessment Measures from the 2011-2012 School Year Predicting Probability of Reading Success from FAIR-K from the 2012-2013 Kindergarten Year***

Predictor Variables	Overall R^2	Unique Variance Accounted for in Model				*N*
		Print Knowledge	Phonological Awareness	Oral Language	Shared Variance	
AP1	.20	.04	.00	.03	.13	30,234
AP2	.30	.10	.01	.01	.18	28,331
AP3	.31	.10	.03	.00	.18	27,603

Notes. R^2 = Variance accounted for by model. AP = Assessment Period. All total and unique variance components were significant at $p < .001$.

Examining Differences in Early Academic Skills Acquisition between Data from 2011-2012 VPK Year and 2012-2013 VPK Year

Using data from both the 2011-2012 and the 2012-2013 VPK year, we examined how children's VPK scores changed throughout the year. Regression analyses were used to determine how children's VPK scores at AP1 relates to their VPK scores at AP3 and how of different characteristics of the children impacted VPK AP3 scores. Table 8 shows a summary of these analyses for the 2011-2012 VPK Year data and Table 9 show a summary of these analyses for the 2012-2013 VPK Year data.

As can be seen in Table 8 (Model 1), in the 2011-2012 VPK year data, children's AP1 VPK assessment scores significantly contributed to their AP3 VPK scores. This result indicates that AP1 scores significantly predict AP3 scores and further supports the cross-time validity of the VPK assessment.

Additional analyses (Models 2 - 5) were used to examine how children's gender, ethnicity, free and reduced lunch classification (FRPL), limited English proficiency status (LEP), and the school in which they attended VPK impacted their VPK AP3 assessment scores, while controlling for VPK AP1 assessment scores. These analyses were used to examine the unique contributions of each of these child characteristics on AP3 VPK scores above and beyond children's initial early academic skills (i.e., AP1 VPK scores).

In Model 2, children's gender and ethnicity were examined as predictors of VPK AP3 assessment scores. Gender significantly contributed to AP3 scores, while controlling ethnicity and AP1 scores, but ethnicity did not significantly contribute to AP3 scores. Although results of

these analyses indicate that girls scored higher on the VPK assessment at AP3 than did boys, gender contributed less than 1% of the variance to AP3 VPK scores in this model.

Table 8. ***Regression Models Predicting VPK AP3 Assessment Scores Using Data from the 2011-2012 School Year***

Predictors in Models	b	SE b	β	*p*	sr^2	R^2	*N*
Model 1						.42	24,384
Assessment Period 1	.52	.00	.65	.001	.42		
Model 2						.42	24,384
Assessment Period 1	.50	.00	.65	.001	.40		
Gender	.60	.10	.03	.001	.00		
Ethnicity	-.01	.02	.00	.51	.00		
Model 3						.42	23,220
Assessment Period 1	.50	.00	.63	.001	.37		
Free/Reduced Lunch	-.11	.11	-.04	.001	.00		
Model 4						.42	23,230
Assessment Period 1	.49	.00	.63	.001	.37		
Limited English Proficiency	-1.08	.15	-.04	.001	.00		
Model 5						.42	32,553
Assessment Period 1	.51	.00	.65	.001	.41		
School	.00	.00	.04	.001	.00		

Note. b = Unstandardized regression coefficient. SE = Standard Error. *β* = Standardized regression coefficient. *sr* = Semi-partial correlation. R^2 = Variance accounted for by model.

In Model 3, we examined how children's FRPL classification contributed to to their AP3 VPK scores in Model 3. As can be seen in Table 8, children's FRPL status did not significantly contribute to VPK assessment scores at AP3.

In Model 4, we examined how children's English proficiency impacted VPK AP3 assessment scores in Model 4. Results indicated that LEP status significantly contributed to AP3 VPK scores. Although this result indicates that having limited English skills leads to lower AP3 VPK scores, LEP status contributed less than 1% of the variance to AP3 VPK scores in this model.

Similarly, in Model 5, the school children attended significantly contributed to AP3 VPK scores, but school contributed less than 1% of the variance to these scores.

Overall, results of these analyses indicate that whereas some characteristics of children were significant predictors of AP3 VPK scores (i.e., gender, FRPL status, LEP status, school attended) none of these characteristics contributed a meaningful amount of variance to children's AP3 VPK scores. Such results indicate that the VPK assessment is a fair measure of early academic skills, measuring these skills similarly in many different types of children. Moreover, that these results were obtained using the 2011-2012 VPK year data indicates that teachers across the state are administering the test in a fair and unbiased manner overall.

Result from the multiple regression analyses of the 2012-2013 VPK year data were similar. As can be seen in Table 9, in the 2012-2013 VPK year, children's AP1 VPK scores significantly contributed to their AP3 VPK scores, supporting the cross-time validity of the VPK assessment in the 2012-2013 year when administration of the VPK assessment became mandatory as a part of accountability testing.

Fewer socio-economic variables were available for this data set because the FLKRS data for this cohort of children were not available; therefore, only gender and the school children attended for VPK were examined as predictors of AP3 VPK scores. Gender significantly contributed to AP3 VPK scores. Although these results indicate that girls scored higher on the AP3 VPK than did boys, gender contributed less than 1% of the variance to AP3 VPK scores in this model. Similarly, the school children attended for VPK significantly contributed to AP3 VPK scores, but it contributed less than 1% of the variance to these scores. As with the 2011-2012 data, whereas these child characteristics contributed significantly and uniquely to AP3 VPK scores, they accounted for trivial amounts of the variation in these scores. In general, these results indicate that the VPK assessment continued to be a fair measure of early academic skills after becoming mandated and part of accountability testing.

Table 9. ***Regression Models Predicting VPK AP3 Assessment Scores Using Data from the 2012-2013 School Year***

Predictors in Models	b	SE b	β	*p*	sr^2	*F*	R^2	*N*
Model 1							.46	125,350
Assessment Period 1	.48	.00	.68	.001	.46			
Model 2							.46	125,350
Assessment Period 1	.48	.00	.67	.001	.45			
Gender	.60	.04	.03	.001	.00			
Model 3							.46	125,350
Assessment Period 1	.48	.00	.68	.001	.46			
School	.00	.00	-.01	.022	.00			

Note. b = Unstandardized regression coefficient. SE = Standard Error. *β* = Standardized regression coefficient. *sr* = Semi-partial correlation. R^2 = Variance accounted for by model.

Relative Change in Children's VPK Scores Across the Preschool Year in 2011-2012 and 2012-2013 VPK Years

It is possible that in the 2011-2012 school year, when the VPK assessment was taken to full scale, teachers at different schools may have administered the VPK assessment inaccurately. To examine this possibility we examined the percentage of children at each school whose VPK assessment scores changed substantially from AP1 to AP3. We first calculated absolute difference scores in the VPK assessments by subtracting children's VPK scores at AP1 from their scores at AP3. A mean absolute difference score across all the participants was computed and an examination of the distribution of these scores conducted. Children whose difference score was one standard deviation above or below the mean difference score were flagged.

On the Print Knowledge subtest, 22% of children had difference scores greater than one standard deviation below the mean and 22% of children had difference scores greater than one standard deviation above the mean. On the Phonological Awareness subtest, 14% of children had difference scores greater than one standard deviation below the mean and 19% of children had difference scores greater than one standard deviation above the mean. On the Oral Language subtest, 10% of children had difference scores greater than one standard deviation below the mean and 15% of children had difference scores greater than one standard deviation above the mean. On the Math subtest, 16% of children had difference scores greater than one standard deviation below the mean and 13% of children had difference scores greater than one standard deviation above the mean. These children's scores on the VPK assessment either improved more than would have been expected or improved less than would have been expected over the course of the year. A non-trivial percentage of children with difference scores more than one standard deviation below the mean scored lower on the VPK at AP3 than they scored at AP1. Specifically, 4, 6, 5, and 4% of children scored lower on the VPK at AP3 than they scored at AP1 on the Print Knowledge, Phonological Awareness, Oral Language, and Mathematics subtests, respectively. Overall, however, the percentage of children scoring more than a standard deviation above or below the mean was close to the percentage that would be expected based on a normal distribution of difference scores (i.e., ~16%).

We then examined preschools in which more than 50% of the children from that school had difference scores greater than one standard deviation above the mean and preschools in which more than 50% of the children from that school had difference scores greater than one standard deviation below the mean. As can be seen in upper panel of Table 10, out of 1,176 schools that reported VPK assessment scores in the 2011-2012 school year, less than 6.0% of schools had more than 50% of children whose scores increased more than would be expected for all of the VPK subtests. Furthermore, in this school year less than 7.0% of schools had more than 50% of students whose scores did not increase as much as would be expected for all subtests. These schools may not have adequately captured the change in early academic skills across the preschool year.

Table 10. ***Numbers and percentages of VPK providers with more than 50% of students having difference scores +/- 1 SD from the mean difference score in 2011-2012 and 2012-2013 VPK year***

	PK		PA		OL		Math	
	N	Percent	N	Percent	N	Percent	N	Percent
2011-2012								
Higher	68	5.78	68	5.78	47	4.00	30	2.55
Lower	77	6.55	47	4.00	20	1.70	58	4.93
2012-2013								
Higher	374	6.25	343	5.74	271	4.53	368	6.15
Lower	368	6.15	159	2.66	56	0.94	199	3.33

Note. 2011-2012 *N* = 1,176. 2012-2013 *N* = 5,980. PK = Print Knowledge. PA = Phonological Awareness. OL = Oral Language. Higher = Number and percent of schools in which greater than 50% of children had absolute difference scores greater than 1 standard deviation above the mean of the average absolute difference score. Lower = Number and percent of schools in which greater than 50% of children had absolute difference scores greater than 1 standard deviation below the mean of the average absolute difference score.

We also examined the possibility that in the 2012-2013 VPK year, when the VPK assessment was mandated and part of accountability testing, teachers at certain schools did not accurately administer the VPK assessment across the school year. Again we examined the percentage of children at each school whose VPK assessment scores changed substantially from AP1 to AP3.

On the Print Knowledge subtest, 23% of children had difference scores greater than one standard deviation below the mean and 21% of children had difference scores greater than one standard above the mean. On the Phonological Awareness subtest, 13% of children had difference scores greater than one standard deviation below the mean and 20% of children had difference scores greater than one standard above the mean. On the Oral Language subtest, 9% of children had difference scores greater than one standard deviation below the mean and 16% of children had difference scores greater than one standard below the mean. On the Mathematics subtest, 16% of children had difference scores greater than one standard deviation below the mean and 19% of children had difference scores greater than one standard below the mean. As was the case in the 2011-2012 VPK year, many children in the 2012-2013 VPK year scored lower at AP3 than they scored at AP1 on the various subtests. Specifically, 3, 6, 4, and 3% of children scored lower at AP3 than they scored at AP1 on the Print Knowledge, Phonological Awareness, Oral Language, and Mathematics subtests, respectively. Overall, however, the percentage of children scoring

more than a standard deviation above or below the mean was close to the percentage that would be expected based on a normal distribution of difference scores (i.e., ~16%).

As with the 2011-2012 data, we then examined preschools in which more than 50% of the children from that school had high absolute difference scores as well as preschools in which more than 50% of the children from that school had very low absolute difference scores. As can be seen in the lower panel of Table 10, out of the 5,980 schools that reported VPK assessment scores in the 2012-2013 school year, less than 7% of schools had more than 50% of children whose scores increased more than would be expected for all of the VPK subtests. Furthermore, in this school year less than 7% of schools had more than 50% of students whose scores did not increase as much as would be expected for all subtests. These schools may not have adequately captured the change in early academic skills across the preschool year.

With the exception of VPK math scores, the percentage of providers with a high percentage of children whose scores increased more than would be expected from AP1 to AP3 was similar in the 2011-2012 data and the 2012-2013 data. The percentage of providers with a high percentage of children whose scores increased less than would be expected from AP1 to AP3 was similar in 2011-2012 and 2012-2013 for the VPK Print Knowledge subtest, but was lower in the 2012-2013 data for the three other subtests than it was in the 2011-2012 data. These results may indicate that some providers in the 2012-2013 VPK year--when use of the VPK measures was mandatory--obtained less accurate scores for children whose scores were lower than average. This pattern would result in higher readiness rates if readiness rates were determined on the basis of AP3 VPK scores. Alternatively, it may be that these providers provide instructional services that promote more growth of children pre-academic skills. Regardless, the percentages of children with growth higher growth than expected was not substantially different across the voluntary- versus mandatory-usage years, suggesting that few providers are reporting inaccurate scores to demonstrate higher rates of growth than actually obtained.

Examining Differences in Growth of Early Academic Skills (2011-2012 and 2012-2013 data)

Using data from both the 2011-2012 and 2012-2013 academic years, we constructed latent growth curve models to examine the trajectories of growth of early academic skills from AP1 to AP3. Because only three time points were available for each of the subtests of the VPK assessment, linear growth models were constructed. Additionally, we used the available child-level variables in each year's data (e.g., gender, age, FRPL status, LEP status, and exceptional student education [ESE] status) as predictors of the intercept (i.e., skill level at AP1) and slope (i.e., rate of growth from AP1 to AP3) in the growth models. When using these variables to predict slope, we controlled for the effect of intercept on slope (i.e., the degree to which initial scores predicted rate of growth).

Model fit statistics and mean intercept and slope parameters from unconditional models (i.e., models with no predictors of intercept and slope) for data from the 2011-2012 academic year are shown in Table 11. Five separate indices of model fit are reported. Typically, lower, non-significant Chi-Square values are indicative of good model fit; however, this test can become

overly sensitive to statistical significance with large sample sizes, such as those in the VPK data. Therefore, other model fit statistics, including the root mean-square error of approximation (RMSEA), comparative fit index (CFI), Tucker-Lewis index (TLI), and square root-mean residual (SRMR) were examined. RMSEA and SRMR values less than .05 and CFI and TLI values close to 1.00 are indicative of good model fit. Based on these statistics, all models provided good fit to the data, indicating that, on average, children's scores on the all of the VPK subtests were increasing across the school year.

Mean intercept and slope values are also shown in Table 11 (next page). Mean intercept values for a given subtest represent the average score on that subtest at AP1. Mean slope values for a given subtest represent the average increase in scores on that subtest between assessment points. For example, an intercept mean of 7.00 and a slope mean of 2.00 would correspond to average scores of 7.00 at AP1, 9.00 at AP2, and 11.00 at AP3.

Parameters for child characteristics predicting intercept in conditional growth curve models for the 2011-2012 academic year are shown in the upper panel of Table 12. To interpret these tables, parameters predicting intercept and slope are added to the average values of intercept and slope. For example, children's average starting phonological awareness score was 7.95 (average intercept for phonological awareness in the conditional model as shown in Table 11). The average child who was female, Hispanic, and LEP would be expected to score 6.36 on the phonological awareness assessment (average intercept plus parameters for gender [.36], Hispanic [-.27], and LEP [-1.68]). The average growth rate in phonological awareness scores is 2.06. Therefore, this child would be expected to have growth across the VPK year of 1.54 (average slope plus parameters for intercept [-.24], gender [.05], Hispanic [-.01], and LEP [-.32]).

All child characteristic variables significantly predicted intercept (i.e., skill level at AP1) for all early academic skills. For all subtests, gender was associated with initial status, such that girls had higher initial VPK scores than did boys. Age was also significantly associated with initial status, indicating that older children had higher initial VPK scores than did younger children. Children who qualified for FRPL, children who were classified as LEP, and children identified as ESE had significantly lower initial VPK scores than did children who were not so classified. Black/African American children and Hispanic/Latino children had weaker initial phonological awareness skills than did children of other races/ethnicities. White children had stronger initial phonological awareness and oral language skills than did children of other races/ethnicities. However, white children had lower initial print knowledge and mathematics skills than did children of other races/ethnicities.

Parameters for child characteristics predicting slope in conditional growth curve models for the 2011-2012 academic year are shown in the lower panel of Table 12. For all subtests, the effect of intercept was significant, such that children who had higher initial VPK scores had slower rates of growth in VPK scores across the school year. Additionally, older children and children classified as ESE at AP1 had significantly slower rates of growth than did children who were younger or classified as ESE. Gender significantly predicted slope such that girls had faster rates of growth of phonological awareness, print knowledge, and oral language skills than did boys. Children who qualified for FRPL had significantly slower rates of growth of print knowledge and oral language skills than did children who did not qualify for FRPL. Children who were

classified as LEP had slower rates of growth of phonological awareness and oral language skills than did children who were not classified as LEP. White children had significantly slower rates of growth of print knowledge skills than did children of other races/ethnicities. Hispanic/Latino children had significantly faster rates of growth of mathematics skills than did children of other races/ethnicities.

Model fit statistics and mean intercept and slope for data from the 2012-2013 academic year are shown in Table 13. All models provided adequate to good fit to the data, indicating that children's scores on the all of the VPK subtests were increasing across the school year. Mean intercept and slope values for all models are also reported in Table 13.

Parameters for child characteristics predicting intercept for data from the 2012-2013 academic year are shown in the upper panel of Table 14. Because data from the 2013-2014 kindergarten year were unavailable, only child age and gender were available as child-level predictors of intercept and slope. Gender was a significant predictor of intercept for all VPK subtests, such that girls had higher initial VPK scores than did boys. Additionally, child age was a significant predictor of intercept, indicating that older children had higher initial VPK scores than did younger children.

Parameters for child characteristics predicting slope for data from the 2012-2013 academic year are shown in the lower panel of Table 14. For all subtests, the effect of intercept was significant, such that children who had higher initial VPK scores had slower rates of growth in VPK scores across the school year. Gender was a significant predictor of slope for the Phonological Awareness, Print Knowledge, and Oral Language subtests of the VPK, indicating that girls' VPK scores on these subtests grew at a faster rate than did boys' VPK scores on these subtests; however, rate of growth of mathematics skills did not differ for girls and boys. Age predicted rate of growth for scores on the Phonological Awareness and Print Knowledge subtests, indicating that phonological awareness and print knowledge skills increased at a faster rate for older children than they did for younger children; however, mathematics skills increased at a faster rate for younger children than they did for older children. Age did not significantly predict rate of growth of oral language skills.

Table 11. ***Fit statistics and mean intercept and slope for 2011-2012 growth curve analyses for pre-K early language, literacy, and math skills***

Model	Chi-Square	RMSEA	CFI	TLI	SRMR	Intercept	Slope
PA – Unconditional	69.34***	.05	.98	.94	.03	7.13	2.11
PA – Conditional	189.29***	.03	.98	.95	.01	7.95	2.06
PK – Unconditional	3.44	.01	1.00	1.00	.02	7.03	2.42
PK – Conditional	38.42***	.01	1.00	.99	.01	8.58	1.77
OL – Unconditional	11.45***	.02	.98	.94	.02	15.86	3.27
OL – Conditional	95.97***	.02	.98	.94	.01	16.84	3.01
Math – Unconditional	207.52***	.06	.93	.90	.04	9.50	2.54
Math – Conditional	490.22***	.04	.96	.90	.02	11.18	2.28

Note. PA = Phonological Awareness; PK = Print Knowledge; OL = Oral Language; RMSEA = Root Mean Square Error of Approximation; CFI = Comparative Fit Index; TLI = Tucker/Lewis Fit Index; SRMR = Standardized Root Mean Square Residual; Intercept = Mean Intercept. Slope = Mean Slope. For PK, T3 freely estimated and residual variance of T1 and T3 fixed to 0. For OL, T3 freely estimated and residual variance of T1 fixed to 0. For Math, T3 residual variance fixed to 0.
*** $p < .001$.

Table 12. ***Parameters from 2011-2012 conditional growth curve models of early language, literacy, and math skills***

	PA	PK	OL	Math
		Intercept		
Gender	.36***	.47***	.48***	.68***
Age	.21***	.20***	.25***	.32***
FRPL	-.86***	-1.70***	-1.31***	-1.86***
LEP	-1.68***	-.98***	-3.41***	-1.71***
ESE	-1.59***	-.92***	-2.04***	-2.06***
White	.45***	-.70***	.93***	-.47***
Black	-1.00***	-.43***	-1.42***	-.74***
Hispanic	-.27*	-.86***	-.43**	-.74***
		Slope		
Gender	.05**	.19***	.14***	-.02
Age	-.01**	-.01**	-.02**	-.04***
FRPL	-.04	-.06*	-.14**	.05
LEP	-.32***	-.03	-.53***	-.03
ESE	-.38***	-.32***	-.57***	-.46***
White	-.02	-.11*	.03	-.06
Black	-.03	.01	-.06	-.03
Hispanic	-.01	.05	.11	.11*
Intercept	-.24***	-.46***	-.46***	-.21***

Note. PA = Phonological Awareness. PK = Print Knowledge. OL = Oral Language. FRPL = Free/Reduced Price Lunch. LEP = Limited English Proficiency. ESE = Exceptional Student Education

*** $p < .001$. ** $p < .01$. * $p < .05$.

Table 13. ***Fit statistics for 2012-2013 growth curve analyses for pre-K early language, literacy, and math skills***

	Chi-Square	RMSEA	CFI	TLI	SRMR	Intercept	Slope
PA – Unconditional	660.21***	.04	.99	.99	.03	7.07	2.15
PA – Conditional	1031.38***	.04	.99	.97	.02	6.98	2.17
PK – Unconditional	630.67***	.04	.99	.98	.05	6.99	2.52
PK – Conditional	788.78***	.03	.99	.98	.03	6.87	2.55
OL – Unconditional	120.54***	.03	1.00	1.00	.03	15.78	3.38
OL – Conditional	192.25***	.02	1.00	.99	.02	15.63	3.45
Math – Unconditional	3159.06***	.10	.96	.94	.04	9.43	3.32
Math – Conditional	4561.95***	.08	.95	.89	.03	9.38	2.63

Note. PA = Phonological Awareness; PK = Print Knowledge; OL = Oral Language; RMSEA = Root Mean Square Error of Approximation; CFI = Comparative Fit Index; TLI = Tucker/Lewis Fit Index; SRMR = Standardized Root Mean Square Residual; Intercept = Mean Intercept. Slope = Mean Slope. For PA, T3 residual variance fixed to 0. For PK, T3 freely estimated and T1 and T3 residual variance fixed to 0. For OL, T3 freely estimated and T1 residual variance fixed to 0. For Math, T3 residual variance fixed to 0.

*** $p < .001$.

Table 14. ***Parameters from 2012-2013 conditional growth curve models of early language, literacy, and math skills.***

	PA	PK	OL	Math
		Intercept		
Gender	.31***	.40***	.50***	.62***
Age	.19***	.19***	.25***	.31***
		Slope		
Gender	.07***	.20***	.12***	.00
Age	.01***	.01***	.00	-.01***
Intercept	-.25***	-.47***	-.44***	-.21***

Note. PA = Phonological Awareness. PK = Print Knowledge. OL = Oral Language. Age = age at time of initial assessment.
*** $p < .001$.

Comparison of Growth Rates Across 2011-2012 and 2012-2013 VPK Years

We also evaluated growth in academic skills across the school year for those VPK providers for which data were available in both the 2011-2012 and the 2012-2013 academic years. Model fit statistics for growth curve models that included only those VPK providers for which data was available for both the 2011-2012 and the 2012-2013 academic years are reported in Table 15. All models provided good fit to the data. Additionally, mean intercept and slope values for both the 2011-2012 and the 2012-2013 academic years are reported in Table 15. Mean intercept and slope values were similar across years, indicating that on average, children in the 2011-2012 and 2012-2013 academic years began the school year with similar skills and grew at similar rates over the course of the school year.

Parameters for child characteristics predicting intercept are reported in the upper panel of Table 16. For all VPK subtests gender was a significant predictor of intercept, indicating that girls began the school year with greater levels of academic skills than did boys. Additionally, age was a significant predictor of intercept for all VPK subtests, indicating that older children had higher initial VPK scores than did younger children. Parameters were similar in magnitude and direction across years.

Parameters for child characteristics predicting slope are reported in the lower panel of Table 16. In both the 2011-2012 and 2012-2013 years the effect of intercept was significant, indicating that children who had higher initial VPK scores had slower rates of growth in VPK scores across the

Table 15. ***Fit statistics and parameter values for VPK providers that provided data in both years***

	Chi-Square	RMSEA	CFI	TLI	SRMR	2011-2012 Intercept	2011-2012 Slope	2012-2013 Intercept	2012-2013 Slope
PA – Unconditional	329.75***	.05	.98	.96	.03	6.99	2.10	7.01	2.25
PA – Conditional	581.92***	.05	.98	.95	.02	6.60	2.12	6.60	2.26
PK – Unconditional	42.02***	.01	.99	.99	.04	6.92	2.46	6.80	2.64
PK – Conditional	76.62***	.01	.99	.99	.03	6.36	2.50	6.37	2.61
OL – Unconditional	40.14***	.02	.99	.97	.03	15.63	3.29	15.59	3.50
OL – Conditional	81.77***	.02	.99	.97	.02	15.03	3.40	15.01	3.62
Math – Unconditional	25.75***	.01	1.00	1.00	.01	9.17	3.21	9.18	3.38
Math – Conditional	87.23***	.02	1.00	.99	.01	8.31	3.47	8.43	3.62

Note. PA = Phonological Awareness; PK = Print Knowledge; OL = Oral Language; RMSEA = Root Mean Square Error of Approximation; CFI = Comparative Fit Index; TLI = Tucker/Lewis Fit Index; SRMR = Standardized Root Mean Square Residual; Y1 = 2011-2012 School Year. Y2 = 2012-2013 School Year. Intercept = Mean Intercept. Slope = Mean Slope. For PA, T3 residual variance fixed to 0 for children in the 2012-2013 school year. For PK, T3 freely estimated and residual variance of T1 and T3 fixed to 0. For OL, T3 freely estimated and residual variance of T1 fixed to 0. For math, T3 freely estimated, residual variance of T3 fixed to 0, and residual variance of T1 fixed to 0 for children in the 2011-2012 school year.

*** $p < .001$

Table 16. ***Unstandardized parameters from growth curve models using only VPK providers that provided data in both years.***

		PA	PK	OL	Math
			Intercept		
2011-2012	Gender	.33***	.43***	.49***	.68***
	Age	.19***	.19***	.23***	.31***
2012-2013	Gender	.31***	.36***	.47***	.58***
	Age	.19***	.19***	.24***	.30***
			Slope		
2011-2012	Gender	.08***	.20***	.14***	.06**
	Age	-.01***	-.01***	-.03***	-.03***
	Intercept	-.23***	-.45***	-.42***	-.31***
2012-2013	Gender	.08***	.21***	.13***	.06**
	Age	.01***	.01***	.00	.01**
	Intercept	-.25***	-.47***	-.43***	-.31***

Note. PA = Phonological Awareness. PK = Print Knowledge. OL = Oral Language. Age = Age at time of initial assessment.
*** $p < .001$; ** $p < .01$

school year. Across the 2011-2012 and 2012-2013 academic years, gender was a significant predictor of slope, indicating that academic skills of girls increased at a faster rate than did the academic skills of boys. For the 2011-2012 academic year, age was a significant, negative predictor of slope, indicating that the academic skills of younger children increased at a faster rate than did the academic skills of older children; however, for the 2012-2013 academic year, age was a significant, positive predictor of slope for the Phonological Awareness, Print Knowledge, and Mathematics subtests, indicating that the phonological awareness, print knowledge, and mathematics skills of older children increased at a faster rate than did the phonological awareness print knowledge and mathematics skills of younger children. Age did not significantly predict rate of growth of children's oral language skills.

Examining Differences in Early Academic Skills Acquisition Across the 2011-2012 versus 2012-2013 VPK Years

To determine whether administration of the VPK measure differed across the 2011-2012 and the 2012-2013 VPK years, multiple regression analyses were conducted using data from VPK providers that provided data in both years. Specifically, scores at AP3 were predicted from scores at AP1, gender, and child age. Results of these analyses are shown in Table 17. Children's scores at AP1 significantly predicted scores at AP3 for all VPK subtests, indicating that children who had higher scores at AP1 also had higher scores at AP3. Across the 2011-2012 and 2012-2013 school years, unstandardized regression coefficients were similar, indicating that administration of the VPK assessment at AP1 and AP3 did not differ across the two school years.

As with the growth models presented above, child gender significantly predicted VPK scores at AP3 across both school years, indicating that girls had higher scores at AP3 on all VPK subtests than did boys. For data from the 2011-2012 VPK year, child age significantly predicted print knowledge, oral language, and math scores at AP3, such that younger children had higher scores at AP3 than did older children. Child age did not significantly predict phonological awareness scores at AP3. For data from the 2012-2013 VPK year, child age significantly predicted phonological awareness, print knowledge, and math scores at AP3, such that older children had higher scores at AP3 than did younger children. Child age did not significantly predict oral language scores at AP3.

Table 17. ***Unstandardized regression coefficients from regressions predicting Time 3 academic outcomes for children in both samples***

		PA	PK	OL	Math
		Time 3			
2011-2012	Time 1 Score	.39***	.31***	.42***	.47***
	Gender	.19***	.27***	.22***	.09*
	Age	-.01	-.02***	-.04***	-.05***
2012-2013	Time 1 Score	.36***	.31***	.43***	.45***
	Gender	.18***	.28***	.17***	.09***
	Age	.05***	.01***	.00	.03***

Note. PA = Phonological Awareness. PK = Print Knowledge. OL = Oral Language. Age = Age at time of initial assessment.
*** $p < .001$; ** $p < 01$; * $p < .05$.

To further examine differences in academic skills development across the 2011-2012 and 2012-2013 school years, we computed absolute difference scores from AP1 to AP3 (i.e., how much children's academic skills grew over the course of the year) and compared these scores across years. These comparisons are shown in Table 18 (next page). Absolute difference scores were significantly different across years for phonological awareness and oral language, but not for print knowledge and math. Specifically, growth in phonological awareness and oral language was greater in the 2012-2013 school year than it was in the 2011-2012 school year. However, effect sizes for these significant differences were small, indicating that statistical significance may have been an artifact of large sample size.

Additionally, mean scores at all assessment points were compared across the 2011-2012 and the 2012-2013 school years for each VPK subtest. Mean scores for print knowledge and math were different across years at all assessment points. Furthermore, mean phonological awareness scores differed across years at AP2 and AP3 and mean oral language skills were different across years at AP3. However, effect sizes were very small, indicating that mean scores differed by less than one tenth of a standard deviation for all comparisons.

Absolute difference scores were computed and compared across years for those VPK providers for which data were available in both the 2011-2012 and the 2012-2013 school years. Analyses indicated that absolute difference scores were significantly different across years for all VPK subtests, such that mean growth in all skills was greater in the 2012-2013 school year than it was in the 2011-2012 school year. However, effect sizes were very small, never exceeding one tenth of a standard deviation difference in absolute difference scores across years.

Mean scores at all assessment points were compared across years for those VPK providers for which data were available in both years. Mean scores for phonological awareness and math were different across years at all assessment points. Furthermore, mean scores for print knowledge and phonological awareness at AP1 and AP3 were significantly different across years. Consistent with analyses of absolute difference scores, effect sizes were very small, never exceeding seven one hundredths of a standard deviation. These findings suggest that administration of the VPK assessment did not differ across the 2011-2012 and 2012-2013 school years.

Table 18. ***Comparisons of scores at each assessment period and absolute difference scores across years.***

Outcome	Y1 *N*	Y1 *M*	Y2 *SD*	Y2 *N*	Y2 *M*	Y2 *SD*	*t*	*d*
PK AP 1	29075	6.98	3.67	157782	7.07	3.69	-3.84***	-.02
PK AP 2	27481	9.46	3.16	119674	9.51	3.13	-2.73**	-.02
PK AP 3	26673	10.52	2.55	137881	10.64	2.41	-6.97***	-.05
PK Diff	23648	3.60	3.31	125362	3.56	3.21	1.58	.01
PA AP 1	29075	7.06	3.56	157775	7.05	3.59	.46	.00
PA AP 2	27481	9.41	3.52	119670	9.55	3.51	-5.97***	-.04
PA AP 3	26673	11.22	3.14	137880	11.40	2.97	-8.39***	-.06
PA Diff	23648	4.31	3.41	125357	4.39	3.42	-3.59***	-.03
OL AP 1	29073	15.89	4.50	157762	15.87	4.56	.62	.01
OL AP 2	27473	19.14	3.65	119659	19.18	3.68	-1.79	-.01
OL AP 3	26666	19.92	3.41	137877	20.12	3.11	-8.92***	-.06
OL Diff	23641	4.12	3.66	125353	4.21	3.47	-3.44**	-.03
Math AP 1	29050	9.28	4.48	157767	9.52	4.53	-8.70***	-.05
Math AP 2	27457	12.41	4.22	119660	12.75	4.18	-11.87***	-.08
Math AP 3	26650	14.45	3.71	137880	14.77	3.48	-13.11***	-.09
Math Diff	23624	5.28	3.73	125357	5.24	3.63	1.45	.01

Note. Y1 = 2011-2012 VPK Year. Y2 = 2012-2013 VPK Year. PK = Print Knowledge, PA = Phonological Awareness, OL = Oral Language, AP = Assessment Point. Diff = Difference Score. ** $p < .01$, *** $p < .001$.

Table 19. ***Comparisons of scores at each assessment period and absolute difference scores for centers that provided data in both years.***

Outcome	Y1 *N*	Y1 *M*	Y2 *SD*	Y2 *N*	Y2 *M*	Y2 *SD*	*t*	*d*
PK AP 1	25575	7.11	3.67	35138	6.89	3.69	7.30***	.06
PK AP 2	24107	9.54	3.11	30352	9.51	3.14	1.20	.01
PK AP 3	23332	10.58	2.50	31101	10.67	2.38	-4.16***	-.04
PK Diff	20694	3.52	3.32	28396	3.81	3.27	-9.70***	-.09
PA AP 1	25576	7.15	3.58	35136	6.95	3.63	6.63***	.06
PA AP 2	24107	9.51	3.49	30351	9.65	3.48	-4.51***	-.04
PA AP 3	23332	11.31	3.08	31101	11.53	2.91	-8.28***	-.07
PA Diff	20695	4.32	3.44	28394	4.64	3.41	-10.18***	-.09
OL AP 1	25574	15.97	4.48	35135	15.69	4.62	7.51***	.06
OL AP 2	24100	19.22	3.57	30347	19.17	3.66	1.43	.01
OL AP 3	23326	19.98	3.36	31101	20.12	3.13	-4.81***	-.04
OL Diff	20689	4.11	3.71	28393	4.38	3.51	-8.34***	-.08
Math AP 1	25551	9.45	4.47	35136	9.29	4.56	4.36***	.04
Math AP 2	24083	12.57	4.16	30347	12.70	4.17	-3.68***	-.03
Math AP 3	23309	14.57	3.61	31101	14.74	3.50	-5.30***	-.05
Math Diff	20671	5.23	3.75	28394	5.47	3.66	-7.05***	-.06

Note. Y1 = 2011-2012 VPK Year. Y2 = 2012-2013 VPK Year. PK = Print Knowledge, PA = Phonological Awareness, OL = Oral Language, AP = Assessment Point. Diff = Difference Score. *** $p < .001$.

Summary and Conclusions

The analyses described in this report were intended to address five questions concerning the psychometric functioning and accurate use of the VPK Assessments when used state wide by VPK providers. Overall, the results of the analyses conducted indicate that the strong psychometric properties in terms of reliability and validity that the VPK Assessment demonstrated in development work and the field trial continue when the assessment is used on a broad scale. Additionally, the analyses provided little evidence that the validity of the VPK assessment varied from the field trial to state-wide use or was affected by the change from voluntary use to mandatory use.

Specifically, results of concurrent and predictive correlational analyses indicate that the VPK Assessment functions similarly to how it functioned in the field trial. Cross-time correlations between subtests of the VPK Assessment (i.e., AP1 - AP2, AP1 - AP3, AP2 - AP3) were similar in magnitude to those obtained in the original field trial, indicating that the specific assessments at the different assessment periods are measuring reliable variance in children's abilities and that differences between children's relative rank in the overall distribution of scores remains somewhat stable from assessment period to assessment period. Similarly, correlations between scores on the subtests of the VPK Assessment at different assessment periods and scores from the FLKRS and FAIR-K, which are administered at the beginning of children's kindergarten year, were similar between the original field trial and the data from the year that the VPK Assessment was used state wide.

Analyses of growth were conducted to determine the overall pattern of changes in children's skills across the VPK year and to determine how different child characteristics influenced rates of growth in oral language, print knowledge, phonological awareness, and math skills during the VPK year. Across both years and on all four skill domains, there was significant growth in scores on the VPK Assessment during the VPK year. The skills children had at the beginning of the VPK year and the rate of growth during the VPK year was significantly affected by each of the child characteristics examined. In general, girls and older children started the VPK year with higher skills and experienced more growth in skills across the VPK year than did boys and younger children. However, this pattern was qualified by the finding that children with higher initial scores demonstrated slower growth overall--likely because they had already been exposed to activities and experiences that fostered the development of age-appropriate skills in oral language, early literacy, and early math.

Data from the 2011-2012 VPK year included more variables indexing child characteristics than did the data from the 2012-2013 VPK year, because those data were linked to information associated with children's scores on the FLKRS. For data from the 2011-2012 VPK year, most child characteristics were predictive of children's initial skills. Children classified as eligible for FRPL, LEP, or ESE started the VPK year with lower skills in all four skill domains than did children without these designations. Additionally, children who were non-white or Latino started the VPK year with lower skills in all four skill domains than did children classified as white. Rates of growth in all four skill areas were most affected by LEP and ESE status. Children designated as LEP or ESE experienced less growth during the VPK year than did children without these designations.

In addition to examining the psychometric characteristics of the VPK Assessment when used state wide and patterns of growth, analyses were conducted to determine the degree to which providers were administering the assessments in a way that yielded accurate scores. Results of several analyses conducted addressed this question. First, there was no evidence that a substantially higher percentage of children were showing more or less growth than would be expected based on the overall pattern of growth for children. For each subtest of the VPK Assessment and across both years, close to the expected 16% of children had scores indicating higher or lower than expected growth. Moreover, children who demonstrated this higher or lower rate of growth were not concentrated in particular VPK providers--as would be expected if an overall pattern of inaccurate administration or reporting of the assessment were occurring. There were fewer providers in the year in which use of the VPK Assessment was mandatory that reported scores indicating that 50% or more of their children made lower than expected progress, when compared to data reported by VPK providers the prior year when use of the VPK Assessment was voluntary. Whether this change was the result of differences in administration procedures for the assessment or differences in instruction cannot be determined.

In analyses of growth of the skills measured by the VPK Assessment, comparisons across the 2011-2012 and 2012-2013 VPK years did not reveal a pattern of differences that would suggest that the change from voluntary use to mandatory use altered the way the measure was administered. That is, levels of initial skills and amount of growth in skills were comparable across years. Direct comparisons of results from VPK providers with data from both years revealed very similar profiles, suggesting a high degree of stability in initial scores and growth from year to year.

Finally, analyses of absolute differences in scores for VPK Assessment subtests at each assessment period and difference in absolute score changes from AP1 to AP3 across the 2011-2012 VPK year and the 2012-2013 VPK year, revealed small year-to-year differences. Although these differences were statistically significant, they represent less than 1/10 of a standard deviation difference and had no consistent pattern in terms of the direction of the difference.

Appendix A: Additional Analyses

Table A1. ***Correlations between subtests of the VPK assessment at AP1, AP2, and AP3 using data from the 2011-2012 VPK year***

VPK Assessment Subtest	Print Knowledge	Phonological Awareness	Oral Language
VPK Assessment Period 1			
Print Knowledge			
Phonological Awareness	.49		
Oral Language	.49	.62	
Math	.70	.60	.62
VPK Assessment Period 2			
Print Knowledge			
Phonological Awareness	.51		
Oral Language	.48	.64	
Math	.68	.62	.57
VPK Assessment Period 3			
Print Knowledge			
Phonological Awareness	.51		
Oral Language	.48	.68	
Math	.68	.65	.61

Notes. $N = 41{,}322$ at AP1, $N = 38{,}339$ at AP2, $N = 37{,}159$ at AP3. All correlations are significant at $p < .001$.

Table A2. ***Correlations between subtests of the VPK assessment at AP1, AP2, and AP3 using data from the 2012-2013 VPK year***

VPK Assessment Subtest	Print Knowledge	Phonological Awareness	Oral Language
VPK Assessment Period 1			
Print Knowledge			
Phonological Awareness	.49		
Oral Language	.49	.61	
Math	.70	.59	.63
VPK Assessment Period 2			
Print Knowledge			
Phonological Awareness	.49		
Oral Language	.46	.62	
Math	.67	.61	.56
VPK Assessment Period 3			
Print Knowledge			
Phonological Awareness	.47		
Oral Language	.40	.63	
Math	.65	.62	.57

Notes. $N = 157{,}760$ at AP1, $N = 119{,}658$ at AP2, $N = 137{,}877$ at AP3. All correlations are significant at $p < .001$.