TECHNICAL ANALYSIS OF THE SLOSSON PHONICS AND STRUCTURAL ANALYSIS TEST

BRADLEY T. ERFORD
JESSICA L. DUTTON
Loyola College in Maryland

Technical characteristics of Slosson Phonics and Structural Analysis Test (SP-SAT) scores were studied using three independent samples of boys and girls aged 6 years, 6 months, through 9 years, 11 months. A decision efficiency study ($n = 100$) resulted in a total predictive value of .86 (sensitivity = .95, specificity = .77, positive predictive power = .80, negative predictive power = .94). Item difficulty was .65 (median), and item discrimination was .52 (median). Interitem consistency was $r = .93$, .93, and .96 for scores on the Phonics Index, Structural Analysis Index, and SP-SAT total standard score, respectively ($n = 375$), whereas 2-week test-retest reliability was $r = .90$, .88, and .90 for scores on the same scales, respectively ($n = 165$). The SP-SAT yielded moderate to high concurrent validity coefficients when compared with the Wide-Range Achievement Test–Third Edition Reading subtest; Woodcock-Johnson: Tests of Achievement–Third Edition Letter-Word Identification, Word Attack, Passage Comprehension, and Writing Samples subtests; and Woodcock-Johnson: Tests of Achievement–Revised Dictation subtest. Practice and research implications are discussed.

Keywords: *reading; phonics; assessment; structural analysis*

The acquisition of phonics skills and phoneme awareness is imperative in the development of reading ability. Phoneme awareness involves separating words into their smallest components and is essential when first learning to read (Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977; Tummer & Nesdale, 1985). Importantly, it has been shown that students who experi-

Bradley T. Erford, Ph.D., is an associate professor of education and Jessica L. Dutton is a graduate student in the school counseling program at Loyola College in Maryland. Please direct all correspondence to Bradley T. Erford, Ph.D., Loyola College in MD, Timonium Graduate Center, 2034 Greenspring Drive, Timonium, MD 21093; phone: 410-617-1509; e-mail: berford@loyola.edu.

Educational and Psychological Measurement, Vol. 65 No. 6, December 2005 1011-1025
DOI: 10.1177/0013164405275665

ence success in the beginning stages of reading and spelling possess high phoneme awareness skills (Blachman, 1984; Calfee, Lindamood, & Lindamood, 1973; Fox & Routh, 1980; Helfgott, 1976; Stanovich, Cunningham, & Cramer, 1984). Yopp (1995) added that phonological processing, the ability to form sounds into phonemes, is the best indicator of reading ability.

In a related literacy process, writing is a way for students to use words in a structural manner. Writing has been shown to play a large role in students' learning of phonics (Asselin, 2001). Research has shown a statistically significant relationship between a student's writing skill and understanding of letter-sound relationships (Dahl & Scharer, 2000). In using knowledge gained from reading in the writing process, students' need for awareness of letters, sounds, and words increases.

Students who have difficulty with reading in first grade have a hard time catching up with peers (Clay, 1991; Juel, 1988; Stanovich, 1986). If students who are at risk for reading problems are not identified early in their academic careers, these problems will most likely follow them in subsequent years (Speece, Mills, Ritchey, & Hillman, 2003).

Both phonemic awareness and structural writing skills are necessary components of students' successful reading achievement. The Slosson Phonics and Structural Analysis Test (SP-SAT) was designed to assess these skills. The SP-SAT is a screening test for students aged 6 to 9 years and is composed of 10 facets of phonics skill development and 10 facets of structural analysis. The SP-SAT has a three-fold purpose: (a) to identify children who are at risk for literacy problems, eligible for remedial programs, or have learning disabilities; (b) to screen children new to a school or agency; and (c) to aid in curriculum evaluation (Perry & Erford, 2003).

The SP-SAT is a 100-item screening test that can be individually administered, scored, and interpreted in approximately 10 to 15 minutes. Each item is scored right/wrong. The SP-SAT was normed on 375 students, aged 6 through 9 years (Grades 1.5 through 4.4), and can be administered by educators or paraprofessionals with some specialized training.

The SP-SAT is composed of two indices, each of which measures 10 facets reflecting emerging literacy activities and conceptual development. Each facet is composed of five items. The Phonics Index measures (a) consonant blends; (b) consonant digraphs; (c) short vowels; (d) long vowels: final *e* and open syllable; (e) long vowels: digraphs; (f) *r*-controlled vowels; (g) *a* followed by *l*, *ll*, *w*, and *u*; (h) dipthongs; (i) hard and soft sounds of *c* and *g*; and (j) silent consonants. The Structural Analysis Index includes (a) indicating root words, (b) indicating suffixes, (c) indicating prefixes, (d) forming plurals, (e) forming possessives, (f) adding suffixes, (g) rewriting contractions, (h) forming compound words, (i) writing abbreviations, and (j) indicating syllables. Each task was selected because it measures reading abilities in the

early elementary years. Early reading skills are a key indicator of future reading proficiency (Reese, 1995; Teale, 1988; Wilson & Thrower, 1985).

It is essential to remember that the SP-SAT is a screening test. A screening test is "a procedure designed to identify children who, because of the risk of a possible learning problem or handicapping condition, should proceed to a more intensive level of diagnostic assessment" (Meisels, 1989, p. 1). In general, screening level tests are held to a lower standard of reliability ($r = .80$ or higher) than diagnostic tests are ($r = .90$ or higher; Salvia & Ysseldyke, 2004) because the decisions stemming from screening tests usually involve whether deeper level, diagnostic testing is warranted.

Purpose

The purposes of this study were to examine the item analysis, internal consistency, test-retest reliability, decision efficiency, and concurrent validity of the SP-SAT. Item analysis involves identifying items that display acceptable item difficulties and item discriminations. Item difficulty in this study was determined using the percentage passing method, whereas item discrimination was determined using the extreme groups method (Anastasi & Urbina, 1997).

Validity involves the usefulness of a test's scores in making accurate decisions in addition to an understanding of what a test really measures and how well it does so (Anastasi & Urbina, 1997). Classical analyses of validity address content, criterion-related, and construct validity. This study assessed only the concurrent validity of the SP-SAT. Evidence of content validity was provided in the SP-SAT manual (Perry & Erford, 2003). Evidence of construct validity of scores was provided through internal consistency and item analysis procedures described above, as well as correlations with other tests purporting to measure similar constructs.

Two methods of criterion-related validity were undertaken. Concurrent criterion-related validity can be established by correlating scores on the SP-SAT and other tests that measure literacy proficiency. Woodcock-Johnson: Tests of Achievement–Third Edition (WJ-III; Woodcock & Johnson, 2001) and Wide-Range Achievement Test–Third Edition (WRAT-3; Wilkinson, 1993) Reading subtest were used in concurrent validity studies.

Decision efficiency is another method of determining a test's scores' criterion-related validity by investigating the accuracy of the SP-SAT in identifying children with literacy difficulties. Procedures discussed by Widiger, Hurt, Frances, Clarkin, and Gilmore (1984) were used to determine the SP-SAT's sensitivity, specificity, positive predictive power (PPP), negative predictive power (NPP), and total predictive value (TPV).

Method

Three independent samples of convenience were used in the following three studies.

Study 1: Item Analysis and Interitem Consistency

PARTICIPANTS

Participants in the standardization sample were 375 students (186 girls, 189 boys) aged 6 ($n = 69$), 7 ($n = 115$), 8 ($n = 138$), and 9 ($n = 53$) years. Of the children, 91.2% were White, 4.3% were African American, 1.1% were Hispanic American, 2.1% were Asian American, 0.5% were Native American, and 0.8% were Other. Students from suburban/urban settings (communities larger than 2,500 people) comprised 78.3% of the sample, whereas 21.7% were from rural settings. Of the participants' fathers, 4.8% had not graduated from high school, 46.9% had graduated high school, 21.1% had completed some college, and 27.2% had graduated from college. Of the participants' mothers, 3.4% had not graduated from high school, 38.1% had graduated high school, 22.4% had completed some college, and 36.1% had graduated from college.

PROCEDURE

SP-SAT protocols were individually administered to participants. Statistical analysis used raw scores to conduct computation of item difficulty, item discrimination (extreme groups method; Anastasi & Urbina, 1997), and interitem consistency.

Study 2: Decision Efficiency

PARTICIPANTS

Participants were 84 children (56 boys, 28 girls) in Grades 1 to 4 (12, 26, 28, and 18 students, respectively) from two schools located in the central Maryland area. Of the students, 72% were White and 28% were African American. Eighty-one percent were from suburban/urban settings (communities larger than 2,500 people), whereas 19% were from rural communities. Of the participants' fathers, 20% had not graduated from high school, 53% had a high school degree, 11% had some college, and 16% had a college degree. Of the participants' mothers, 12% had not graduated from high school, 61% had graduated high school, 18% had completed some college, and 9% had graduated from college.

PROCEDURE

Forty-two children (six 6-year-olds, thirteen 7-year-olds, fourteen 8-year-olds, and nine 9-year-olds) were identified by their teachers as having normal age-appropriate literacy skills, and 42 children (six 6-year-olds, thirteen 7-year-olds, fourteen 8-year-olds, and nine 9-year-olds) were identified by their teachers as having deficient reading skills (based on a combination of a group-administered standardized reading test and global teacher judgment/report). Each teacher had instructed the student in reading for at least 3 months. Each student was administered the SP-SAT by evaluators blind to the student's categorization. Decision efficiency procedures were used to analyze the data. These procedures included computation of sensitivity, specificity, PPP, NPP, and TPV. The cutoff criterion for a decision of deficient reading skills on the SP-SAT was less than the 25th percentile on the total standard score. School systems frequently use this criterion to determine student eligibility for remedial programs.

Study 3: Test-Retest Reliability and Concurrent Validity

PARTICIPANTS

Participants were 165 children (85 girls, 80 boys) in Grades 1 (n = 38), 2 (n = 41), 3 (n = 44), and 4 (n = 42), from two schools located in the central Maryland area. Of the children, 76% were White and 24% were African American. Seventy-eight percent of the children were from suburban/urban settings (communities larger than 2,500 people), whereas 22% were from rural settings. Of the participants' fathers, 4% had not graduated from high school, 39% had graduated high school, 19% had completed some college, and 38% had graduated from college. Of the participants' mothers, 4% had not graduated from high school, 42% had graduated high school, 14% had completed some college, and 40% had graduated from college.

PROCEDURE

The SP-SAT, WRAT-3 Reading subtest, and WJ-III (Dictation, Writing Samples, Word Attack, Letter-Word Identification, and Passage Comprehension) subtests were administered to the participants. The tests were administered in a counterbalanced sequence to eliminate systematic variations due to order of administration. After 2 weeks, all protocols were readministered to the participants. Standard scores for each protocol were used to compute Pearson correlation coefficient computations for test-retest reliability and concurrent validity studies. Correlations were computed for the total sample and for each age category.

INSTRUMENTS

WJ-III Reading subtests. The WJ-III (McGrew & Woodcock, 2001; Woodcock & Johnson, 2000) Reading subtests (Letter-Word Identification, Passage Comprehension, and Word Attack) were used. In addition, the WJ-III Writing Samples subtest and Woodcock-Johnson: Tests of Achievement–Revised (WJ-R; Woodcock & Johnson, 1990) Dictation subtest were used. These subtests were designed to assess basic word recognition and reading comprehension and basic writing skills. Each subtest can be reported individually as a standard score, percentile rank, and grade equivalent. The WJ-III was standardized on 8,818 individuals aged 2 years to adult. The three reading subtests take about 15 to 30 minutes to administer and score. The two writing subtests require an additional 15 to 30 minutes. Median internal consistency coefficients for scores on the Word Identification, Passage Comprehension, Word Attack, Dictation, and Writing Samples subtests for ages 6 to 9 were reported in the manual to be .96, .94, .92, .89, and .88, respectively.

Criterion-related validity studies (Ford, Simmons, & North, as cited in McGrew & Woodcock, 2001; $n = 52$) resulted in statistically significant correlations between the WJ-III Broad Reading Cluster and Kaufman Test of Educational Achievement (Kaufman & Kaufman, 1985) of $r = .76$ with the Reading Composite (RCP) subtest, $r = .67$ with the Reading Decoding subtest, and $r = .65$ with the Reading Comprehension subtest. The WJ-III Broad Reading scores were also correlated with the Wechsler Individual Achievement Test (WIAT; Wechsler, 1992) as $r = .67$ with the RCP scores, $r = .63$ with the Basic Reading subtest, and $r = .68$ with the Reading Comprehension subtest. The WJ-III Written Language scores also correlated as $r = .69$ and .57 with the WIAT Writing Composite and Written Expression subtest, respectively.

Test-retest reliability for scores on the Letter-Word Identification ($n = 453$), Passage Comprehension ($n = 452$), and Word Attack ($n = 435$) subtests yielded results of .95, .92, and .83, respectively, for ages 4 to 17 years. The WRAT-3 manual reported alternate form reliability for scores on the Passage Comprehension subtest only, with the majority of the correlations ranging from .85 to .96.

WRAT-3: Reading subtest. The WRAT-3 Reading subtest (Wilkinson, 1993) is an individually administered test designed to assess basic word recognition skills by having the student orally read a list of words presented in increasing order of difficulty and is scored right/wrong. The Reading subtest score can be reported as a standard score, percentile rank, and grade equivalent. The WRAT-3 was standardized on 4,433 individuals aged 5 years to adult. The Reading subtest takes about 10 minutes to administer and score. Internal consistency coefficients for students aged 6 years, 6 months, to 9 years, 11 months, ranged from .88 to .91, with a median coefficient of .90 for

scores on the "blue" form (n = 1,178). Alternate form reliabilities for scores on the blue and "tan" versions of the Reading subtest (n = 1,178) were reported in the manual to range from .88 to .95 (median r_{ab} = .915). Criterion-related validity studies resulted in correlations of r = 67 (n = 46) between the WRAT-3 Reading subtest and California Test of Basic Skills (fourth edition) Vocabulary subtest, r = .72 (n = 49) between the WRAT-3 Reading subtest and California Achievement Test–Form E, and .80 (n = 31) with the Stanford Achievement Test Vocabulary subtest.

Results and Discussion

Reliability

Interitem consistency of responses to the SP-SAT, computed using the Kuder-Richardson–Formula 20, for the standardization participants (n = 375) was r = .93 (.88, .90, .88, and .87 for ages 6 to 9, respectively) for the Phonics Index, r = .93 (.83, .86, and .80 for ages 7 to 9, respectively) for the Structural Analysis Index, and r = .96 (.93, .90, and .92 for ages 7 to 9, respectively) for the SP-SAT total standard scores. Interitem consistency coefficients for participants in Study 3 are presented in Tables 1 through 5 and are similar to the coefficients derived from the standardization samples' protocols and reported above. The 2-week test-retest reliability (Pearson) coefficients for the participants' responses in Study 3 were r = .90 for the total sample (r = .87, .88, .78, and .85 for Grades 1 to 4, respectively) for the Phonics Index, r = .88 (.70, .81, and .61 for Grades 2 to 4, respectively) for the Structural Analysis Index, and r = .90 (.85, .86, and .91 for Grades 2 to 4, respectively) for the SP-SAT total standard score. Thus, the reliability of SP-SAT scores, both internal consistency and test-retest, was moderately high and generally within the .80+ range of acceptability given the test's screening-level purpose (Salvia & Ysseldyke, 2004).

Item Analysis

The 100 SP-SAT item difficulties, using the percentage passing technique (Anastasi & Urbina, 1997), ranged from .05 to .93, with a median coefficient of .65. Item discriminations ($D = P_H - P_L$; where P_H is the percentage of the top 27% of students who passed the item and P_L is the percentage of the lowest 27% of students who passed the item) using the extreme groups method (Anastasi & Urbina, 1997) ranged from .12 to .81, with a median discrimination of .52. Fifty-four percent of the SP-SAT items met or exceeded the .50 item discrimination criterion. In general, the SP-SAT items presented with good difficulty and discrimination characteristics for a screening-level test.

Table 1
Pearson Correlation Coefficients Between SP-SAT, WRAT-3, and WJ-III Standard Scores for Grade 1

	SPI	WRAT-3	WWI	WWA	WPC	r_{ic}	r_{tt}
First administration							
SPI	—						
WRAT-3	.80	—				.86	
WWI	.80	.87	—			.85	
WWA	.75	.74	.78	—		.85	
WPC	.65	.63	.68	.65	—	.82	
Retest							
SPI	—						.87
WRAT-3	.83	—				.86	.68
WWI	.87	.84	—			.87	.91
WWA	.85	.76	.80	—		.86	.84
WPC	.68	.86	.87	.71	—	.81	.91

Note. $n = 38$. SP-SAT = Slosson Phonics and Structural Analysis Test; WRAT-3 = Wide-Range Achievement Test–Third Edition Reading subtest; WJ-III = Woodcock-Johnson: Tests of Achievement–Third Edition; SPI = SP-SAT Phonics Index; WWI = WJ-III Word Identification subtest; WWA = WJ-III Word Attack subtest; WPC = WJ-III Passage Comprehension subtest; r_{ic} = interitem consistency coefficient; r_{tt} = test-retest reliability coefficient. All correlations $p < .01$.

Decision Efficiency

The proportion of accurately examined students to students recognized by teachers to have an emerging literacy skill deficiency, defined as sensitivity, was .95. The proportion of accurately screened students to students reported by teachers as not having a literacy skill deficiency, defined as specificity, was .77. The proportion of students classified as having literacy skill deficiencies to those students classified as deficient by the SP-SAT, defined as PPP, was .80. And the proportion of students without classification by the teacher as having skill deficiencies to those students classified by the SP-SAT as not deficient, defined as NPP, was .94. The SP-SAT correctly classified 86% of the cases with regard to teacher judgments of literacy nondeficiency and deficiency, giving it a TPV of .86. Table 6 contains the specific decision efficiency data for this study. Although the TPV, specificity, and PPP of the SP-SAT were high for this sample, the SP-SAT had a slight propensity to overidentify students with teacher-identified literacy deficiencies.

False positives are less of a concern in a screening process (Mantzicopoulos & Morrison, 1994). Conversely, it is important for false negatives to be accorded more serious consideration, as such a decision might rule out intervention for students who might otherwise be qualified. The SP-SAT performed satisfactorily in this respect as only two false-negative cases were reported.

(text continues on p. 1023)

Table 2
Pearson Correlation Coefficients Between SP-SAT, WRAT-3, and WJ-III Standard Scores for Grade 2

	STSS	SPI	SSAI	WRAT-3	WWI	WWA	WPC	WDI	WWS	r_{ic}	r_{tt}
First administration											
STSS	—									.92	
SPI	.87	—								.92	
SSAI	.78	.37	—							.87	
WRAT-3	.69	.78	.32	—						.81	
WWI	.59	.63	.32	.82	—					.87	
WWA	.80	.83	.45	.78	.69	—				.89	
WPC	.69	.62	.51	.71	.67	.72	—			.76	
WDI	.65	.50	.59	.62	.52	.62	.63	—		.79	
WWS	.69	.69	.42	.63	.49	.61	.63	.56	—	.82	
Retest											
STSS	—									.93	.85
SPI	.90	—								.92	.88
SSAI	.83	.50	—							.87	.70
WRAT-3	.80	.80	.56	—						.80	.91
WWI	.84	.81	.62	.81	—					.87	.87
WWA	.83	.88	.52	.78	.80	—				.90	.80
WPC	.78	.71	.63	.72	.78	.71	—			.78	.81
WDI	.79	.73	.66	.70	.76	.73	.74	—		.82	.92
WWS	.54	.52	.54	.41	.41	.51	.54	.58	—	.73	.68

Note. $n = 41$. SP-SAT = Slosson Phonics and Structural Analysis Test; WRAT-3 = Wide-Range Achievement Test–Third Edition Reading subtest; WJ-III = Woodcock-Johnson: Tests of Achievement–Third Edition; STSS = SP-SAT Total Standard Score; SPI = SP-SAT Phonics Index; SSAI = SP-SAT Structural Analysis Index; WWI = WJ-III Word Identification subtest; WWA = WJ-III Word Attack subtest; WPC = WJ-III Passage Comprehension subtest; WDI = WJ-R Dictation subtest; WWS = WJ-III Writing Samples subtest; r_{ic} = interitem consistency coefficient; r_{tt} = test-retest reliability coefficient. All correlations $p < .05$.

Table 3
Pearson Correlation Coefficients Between SP-SAT, WRAT-3, and WJ-III Standard Scores for Grade 3

	STSS	SPI	SSAI	WRAT-3	WWI	WWA	WPC	WDI	WWS	r_{ic}	r_{tt}
First administration											
STSS	—									.90	
SPI	.81	—								.85	
SSAI	.86	.40	—							.87	
WRAT-3	.72	.75	.47	—						.66	
WWI	.72	.80	.42	.83	—					.85	
WWA	.81	.83	.54	.81	.83	—				.90	
WPC	.46	.41	.38	.54	.63	.54	—			.70	
WDI	.78	.62	.68	.60	.63	.71	.51	—		.77	
WWS	.57	.34	.59	.29	.41	.38	.27	.48	—	.57	
Retest											
STSS	—									.88	.86
SPI	.75	—								.80	.78
SSAI	.87	.32	—							.88	.81
WRAT-3	.59	.62	.38	—						.79	.81
WWI	.68	.73	.41	.74	—					.87	.90
WWA	.74	.83	.44	.77	.80	—				.90	.88
WPC	.58	.32	.58	.39	.53	.35	—			.68	.77
WDI	.68	.57	.55	.57	.64	.66	.57	—		.77	.72
WWS	.52	.48	.39	.50	.66	.48	.55	.59	—	.67	.73

Note. $n = 44$. SP-SAT = Slosson Phonics and Structural Analysis Test; WRAT-3 = Wide-Range Achievement Test–Third Edition Reading subtest; WJ-III = Woodcock-Johnson: Tests of Achievement–Third Edition; STSS = SP-SAT Total Standard Score; SPI = SP-SAT Phonics Index; SSAI = SP-SAT Structural Analysis Index; WWI = WJ-III Word Identification subtest; WWA = WJ-III Word Attack subtest; WPC = WJ-III Passage Comprehension subtest; WDI = WJ-R Dictation subtest; WWS = WJ-III Writing Samples subtest; r_{ic} = interitem consistency coefficient; r_{tt} = test-retest reliability coefficient. All correlations $p < .05$.

Table 4
Pearson Correlation Coefficients Between SP-SAT, WRAT-3, and WJ-III Standard Scores for Grade 4

	STSS	SPI	SSAI	WRAT-3	WWI	WWA	WPC	WDI	WWS	r_{ic}	r_{tt}
First administration											
STSS	—									.85	
SPI	.89	—								.85	
SSAI	.72	.31	—							.73	
WRAT-3	.79	.78	.46	—						.73	
WWI	.75	.75	.40	.92	—					.91	
WWA	.85	.84	.47	.91	.86	—				.92	
WPC	.74	.65	.54	.79	.74	.72	—			.72	
WDI	.70	.66	.45	.78	.81	.76	.59	—		.74	
WWS	.28	.23	.23	.27	.24	.27	.23	.39	—	.58	
Retest											
STSS	—									.89	.91
SPI	.94	—								.86	.85
SSAI	.86	.63	—							.74	.61
WRAT-3	.76	.79	.53	—						.88	.89
WWI	.85	.85	.66	.86	—					.92	.98
WWA	.81	.85	.57	.87	.91	—				.92	.86
WPC	.66	.77	.35	.71	.74	.78	—			.85	.82
WDI	.68	.68	.51	.77	.84	.70	.58	—		.78	.91
WWS	.31	.33	.22	.39	.49	.37	.44	.44	—	.58	.50

Note. $n = 42$. SP-SAT = Slosson Phonics and Structural Analysis Test; WRAT-3 = Wide-Range Achievement Test–Third Edition Reading subtest; WJ-III = Woodcock-Johnson: Tests of Achievement–Third Edition; STSS = SP-SAT Total Standard Score; SPI = SP-SAT Phonics Index; SSAI = SP-SAT Structural Analysis Index; WWI = WJ-III Word Identification subtest; WWA = WJ-III Word Attack subtest; WPC = WJ-III Passage Comprehension subtest; WDI = WJ-R Dictation subtest; WWS = WJ-III Writing Samples subtest; r_{ic} = interitem consistency coefficient; r_{tt} = test-retest reliability coefficient. All correlations $p < .05$.

Table 5
Pearson Correlation Coefficients Between SP-SAT, WRAT-3, and WJ-III Standard Scores for Total Sample

	STSS	SPI	SSAI	WRAT-3	WWI	WWA	WPC	WDI	WWS	r_{ic}	r_{tt}
First administration											
STSS	—									.93	
SPI	.76	—								.89	
SSAI	.87	.45	—							.93	
WRAT-3	.75	.86	.52	—						.90	
WWI	.76	.81	.60	.90	—					.93	
WWA	.84	.66	.36	.63	.57	—				.92	
WPC	.66	.75	.58	.83	.85	.53	—			.87	
WDI	.80	.49	.81	.67	.74	.52	.66	—		.87	
WWS	.69	.55	.54	.52	.51	.45	.54	.58	—	.77	
Retest											
STSS	—									.93	.90
SPI	.78	—								.88	.90
SSAI	.88	.49	—							.93	.88
WRAT-3	.73	.83	.54	—						.90	.90
WWI	.82	.85	.65	.90	—					.93	.96
WWA	.71	.86	.42	.83	.83	—				.92	.67
WPC	.76	.76	.67	.82	.86	.70	—			.88	.90
WDI	.81	.56	.78	.69	.78	.59	.72	—		.87	.92
WWS	.58	.48	.52	.48	.57	.48	.56	.62	—	.72	.67

Note. $n = 165$. SP-SAT = Slosson Phonics and Structural Analysis Test; WRAT-3 = Wide-Range Achievement Test–Third Edition Reading subtest; WJ-III = Woodcock-Johnson: Tests of Achievement–Third Edition; STSS = SP-SAT Total Standard Score; SPI = SP-SAT Phonics Index; SSAI = SP-SAT Structural Analysis Index; WWI = WJ-III Word Identification subtest; WWA = WJ-III Word Attack subtest; WPC = WJ-III Passage Comprehension subtest; WDI = WJ-R Dictation subtest; WWS = WJ-III Writing Samples subtest; r_{ic} = interitem consistency coefficient; r_{tt} = test-retest reliability coefficient. All correlations $p < .05$.

Table 6
SP-SAT Decision Reliability of 84 Children Classified Normal or Emerging Literacy Skill Deficient by Their Teachers

	Teacher Judgment/Report Decision	
	Deficient	Nondeficient
SP-SAT decision		
Deficient (≤25th percentile)	Valid acceptance (VA) Accurate decision True positives *n* = 39	False positives (FP) Inaccurate decision Overidentifications *n* = 10
Nondeficient (>25th percentile)	*n* = 2 False negatives (FN) Inaccurate decision Underidentifications	*n* = 33 Valid rejections (VR) Accurate decision True negatives

Note. Sensitivity = VA/(VA + FN) = .95; specificity = VR/(VR + FP) = .77; positive predictive power = VA/(VA + FP) = .80; negative predictive power = VR/(VR + FN) = .94; total predictive value = (VA + VR)/(VA + VR + FN + FP) = .86.

Criterion-Related Validity

Criterion-related validation results of the SP-SAT using standard scores of the WRAT-3 Reading subtest and selected WJ-III subtests across two administrations are presented in Tables 1 through 5. On the administration, the SP-SAT total scale displayed a high degree of relationship with the WRAT-3 Reading subtest (r = .75 for the total sample; r = .69, .72, and .79 for Grades 2 to 4, respectively), WJ-III Letter-Word Identification subtest (r = .76 for the total sample; r = .59, .72, and .75 for Grades 2 to 4, respectively), WJ-III Word Attack subtest (r = .84 for the total sample; r = .80, .81, and .85 for Grades 2 to 4, respectively), WJ-III Passage Comprehension subtest (r = .66 for the total sample; r = .69, .46, and .74 for Grades 2 to 4, respectively), WJ-III Writing Samples subtest .69 (r = .69, .57, and .28 for Grades 2 to 4, respectively), and WJ-R Dictation subtest .80 (r = .65, .78, and .70 for Grades 2 to 4, respectively). These coefficients were generally moderate to high, indicating acceptable concurrent validity. The correlations of the second administration of these measures were equivalent in most respects.

Practice and Research Implications

Further research using the SP-SAT is needed to duplicate these studies and explore the reliability and validity of SP-SAT scores on at-risk and special-needs populations. As the relationship between early literacy skills and school-aged reading skills continues to be investigated and understood, pre-

dictive validation studies are of the utmost importance. Such studies will allow educators in the first or second grade to accurately identify students at risk for later reading problems to initiate remedial services.

As mentioned above, the purpose of the SP-SAT is threefold: (a) screening students new to an academic program to ascertain developmental reading and structural analytic levels when no suspicion of difficulty is present, (b) identifying students at risk for reading difficulties when a suspicion of such difficulty exists, and (c) evaluation of a reading curriculum. The samples from this study were mostly in line with this first purpose. The exception was the sample used for the decision efficiency sample, which was more in keeping with the second purpose. In this regard, the SP-SAT performed well (TPV = .86) on the at-risk population under study. Noticeably absent was any evidence that the SP-SAT has usefulness for evaluation of a school's early elementary reading curriculum. This should certainly be explored in future research.

Limitations in this study included the overrepresentation of students with parents who graduated from college as well as fewer students from diverse backgrounds than would typically be expected in a normal population. Still, in concurrence with the SP-SAT's purpose, screening 6- to 9-year-old academically at-risk students to determine a need for further diagnostic testing, the SP-SAT items appeared to yield reliable and valid scores for the studies' samples. The SP-SAT's advantages over other currently available screening devices, such as the WRAT-3 or Test of Early Reading Ability (TERA), include its comprehensiveness within a quick administration and scoring format and high correlations with criterion measures.

References

Asselin, M. (2001). Word recognition. *Teacher Librarian*, *29*, 57-60.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Columbus, OH: Prentice Hall.

Blachman, B. A. (1984). Relationship of rapid naming ability and language analysis skills to kindergarten and first-grade reading achievement. *Journal of Educational Psychology*, *76*, 610-622.

Calfee, R. C., Lindamood, P., & Lindamood, C. (1973). Acoustic-phonetic skills and reading: Kindergarten through twelfth grade. *Journal of Educational Psychology*, *64*, 293-298.

Clay, M. (1991). *Becoming literate: The construction of inner control*. Portsmouth, NH: Heinemann.

Dahl, K. L., & Scharer, P. L. (2000). Phonics teaching and learning in whole language classrooms: New evidence from research. *Reading Teacher*, *53*, 584-595.

Fox, B., & Routh, D. K. (1980). Phonemic analysis and severe reading disability in children. *Journal of Psycholinguistic Research*, *9*, 115-119.

Helfgott, J. (1976). Phoneme segmentation and blending skills of kindergarten children: Implications for beginning reading acquisition. *Contemporary Educational Psychology*, *1*, 157-169.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, *80*, 437-447.

Kaufman, A., & Kaufman, N. (1985). *Kaufman Test of Educational Achievement*. Circle Pines, MN: American Guidance Services.

Liberman, I., Shankweiler, D., Liberman, A., Fowler, C., & Fischer, E. (1977). Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading* (pp. 207-225). Hillsdale, NJ: Lawrence Erlbaum.

Mantzicopoulos, P. Y., & Morrison, D. (1994). Early prediction of reading achievement. *Remedial and Special Education*, *15*, 244-251.

McGrew, K., & Woodcock, R. (2001). *Technical manual for the Woodcock-Johnson: Tests of Achievement–Third Edition*. Itasca, IL: Riverside.

Meisels, S. J. (1989). *Developmental screening in early childhood: A guide* (3rd ed.). Washington, DC: National Association for the Education of Young Children.

Perry, L. A., & Erford, B. T. (2003). *Manual for the Slosson Phonics and Structural Analysis Test (SP-SAT)*. East Aurora, NY: Slosson.

Reese, E. (1995). Predicting children's literacy from mother-child conversations. *Cognitive Development*, *10*, 381-405.

Salvia, J., & Ysseldyke, J. (2004). *Assessment* (9th ed.). Boston: Houghton-Mifflin.

Speece, D. L., Mills, C., Ritchey, K. D., & Hillman, E. (2003). Initial evidence that letter fluency tasks are valid indicators of early reading skill. *Journal of Special Education*, *36*, 223-234.

Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360-406.

Stanovich, K. E., Cunningham, A. E., & Cramer, B. B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology*, *38*, 175-190.

Teale, W. H. (1988). Developmentally appropriate assessment of reading and writing in the early childhood classroom. *Elementary School Journal*, *89*, 173-183.

Tummer, W. E., & Nesdale, A. R. (1985). Phonemic segmentation skill and beginning reading. *Journal of Educational Psychology*, *77*, 417-427.

Wechsler, D. (1992). *Wechsler Individual Achievement Test*. San Antonio, TX: Psychological Corporation.

Widiger, T. A., Hurt, S. W., Frances, A., Clarkin, J. F., & Gilmore, M. (1984). Diagnostic efficiency and *DSM-III*. *Archives of General Psychiatry*, *41*, 1005-1012.

Wilkinson, G. (1993). *Manual for the Wide Range Achievement Test –Third Edition (WRAT-3)*. Wilmington, DE: Jastak Associates.

Wilson, L. C., & Thrower, J. (1985). Early childhood reading educator's perceptions of reading readiness. *Reading Research and Instruction*, *25*, 21-33.

Woodcock, R., & Johnson, R. (1990). *Woodcock-Johnson: Tests of Achievement–Revised*. Itasca, IL: Riverside.

Woodcock, R., & Johnson, R. (2000). *Woodcock-Johnson: Tests of Achievement–Third Edition*. Itasca, IL: Riverside.

Yopp, H. K. (1995). A test for assessing phonemic awareness in young children. *The Reading Teacher*, *49*, 20-29.