CELEBRATING THE *Freedom* OF LITERACY

Editors
Mary Beth Sampson, Patricia E. Linder, Jo Ann R. Dugan, Barrie Brancato

Relating African American Students' Scores on State-Mandated Reading and Writing Tests to Standardized Measures of Reading and Oral and Written Language

Monica Gordon Pershey

Cleveland State University

Abstract

Mandated testing is prominent among educational initiatives. Questions arise about testing students who face economic disadvantages, are of racial, ethnic, or language minority status, or are at risk for academic failure because their schools provide limited opportunities for academic advancement. This study used normed tests to measure oral and written language and reading abilities in 263 grade four and six African American students and compared results to performance on criterion-referenced mandated testing of reading and writing.

For normed and mandated tests, group mean scores were at times below expectations. Principal components analyses determined language capabilities underlying performance on testing. Factors that accounted for the variance in scores included vocabulary, knowledge of the conventions of written language, knowledge of sentence syntax, and reading comprehension.

Prominent among current educational initiatives is the importance of preparing students to meet curriculum standards and perform well on mandated testing (National Education Goals Panel [NEGP], 1994; United States Department of Education [USDOE], 2001; USDOE, 2002). The effort to establish summative tests of academic performance overlaps the current school accountability movement, which holds schools and educators responsible for students' progress (National Center on Educational Outcomes [NCEO], 2001; USDOE, 2002).

Testing figures prominently into the lives of pupils in Ohio. Districts have been required to administer the *Ohio Proficiency Test* (OPT) (Ohio Department of Education, 1995; *Ohio Student Testing Requirements,* 1991) to measure how students perform on learning outcomes identified by the state. The OPT is a criterion-referenced test of content areas and reading and writing given in grades four, six, nine, ten, and twelve. Testing is supposed to correspond with what is taught and measure cumulative learning and overall school achievement. However, preparing students to demonstrate the applied process skills tested by the OPT reading and writing subtests has been a challenge (e.g., summarization, finding paragraph details, writing a friendly letter) (Pottle, 2001).

The OPT has become the standard by which pupil competency and school accountability have been assessed. The state of Ohio ranks school districts according to twenty-seven performance standards, twenty-five of which apply to achieving a 75% pass rate on the subtests of the OPT in all grades. The other two standards are attendance and graduation rates. A district is then assigned to one of four performance categories (i.e, effective, continuous improvement, academic watch, academic emergency). At the time of this study 69 of the state's 607 districts met eight or fewer standards and were in academic emergency (Ohio Department of Education, 2000).

The author of this study provided professional development seminars to teachers from a district that was in academic emergency. Enrollment in this district was virtually 100% African American. The teachers emphasized their impressions that students' consistently poor OPT performance was likely due to an inadequate oral language basis that impacted negatively upon reading and writing performance. Language development, as a factor in test performance, has been suggested by Meisels (1989) and Popham (1999); however, this appears to be a new area of research. Little data exist on how language capabilities correlate with performance on criterion-referenced tests of curriculum mastery. To explore this hypothesized connection, the author collected data on the oral and written language and reading skills of the district's fourth and sixth graders.

This study (a) explored whether students' normed test scores met age and grade level expectations, (b) identified where performance on the normed tests and the OPT were related, and (c) determined language capabilities that predicted OPT reading and writing performance.

Debates Attendant to Testing

Underlying a belief in high-stakes testing is the assumption that students score as they do because of exposure to school curriculum (Glaser, 1994). However, it is problematic to assume that exposure is of uniform quality across

schools or is equally meaningful to all students. Questions arise about the appropriateness of testing students who face economic disadvantages, are of racial, ethnic, or language minority status, or are at risk for academic failure because their schools lack resources and provide limited opportunities for academic advancement (Bobbett, 1993; Denoyer & White, 1992; Gallagher, 1993; International Reading Association [IRA], 1999; Kohn, 2000; Ladson-Billings, 1994; Manning, Lucking, & MacDonald, 1995; McGee, 1997; Meisels, 1989; Popham, 1999; Raivetz, 1992; Roderick, 1995; Westbury, 1994; Winfield, 1990). Studies have shown that minority and low income students are more likely to fail high-stakes tests and remediation for those who fail is less likely to be effective (Morris, 2000; National Assessment of Educational Progress [NAEP], 1998; NAEP, 1999; NAEP, 2000; Schellenberg, 1998; Winfield, 1990). Tests reveal a widening gap between high and low income schools, especially on questions that require an extended verbal response (National Educational Research Policy and Priorities Board, 1999). Summative tests do not identify the academic supports that students would need to receive in order to perform better on curriculum demands and on subsequent testing (Barkley, 2001).

Several studies have shown that minority and low income students are more likely to fail mandated achievement tests and that remediation for those who fail is less likely to be effective (Morris, 2000; NAEP, 1998; NAEP, 1999; NAEP, 2000; Schellenberg, 1998; Winfield, 1990). However, the cause of test failure is not entirely clear. Jackson (1999) offered that risk is less dependent upon income and more directly related to residential and school segregation. Popham (1994,1999) ventured that testing outcomes tend to be biased in favor of children from economically advantaged, stimulus-rich homes, who have acquired language-based information outside of school. A strong language basis, promoted early on in children's education, prepares learners to be developmentally ready to profit from instruction (Meisels, 1989). The language of the school curriculum, including the discourses of literacy instruction, may be beyond the comprehension of children who lack substantial language backgrounds (Butler, 1999; Falk-Ross, 2002; Snow, Burns, & Griffin, 1998).

Characteristics of the OPT

As a criterion-referenced measure, the OPT resembles the NAEP (2000) in some ways. Both reading measures address how well a learner comprehends what is read, extends comprehension, and manipulates information (e.g., follows directions). Both writing assessments require the student to write to accomplish a variety of purposes (Commission on Behavioral and Social Sciences and Education, 1998; NAEP, 2000; Ohio Department of Education, 2000; *Ohio Student Testing Requirements,* 1991).

Despite this resemblance, questions of reliability, validity, and appropriateness arise when education agencies prepare their own criterion-referenced competency tests (Lanese, 1992; Snodgrass & Salzman, 1998). Instruments that are not normed can produce unstable results, particularly when testing African American students (Urdan & Davis, 1998). The *Ohio Statehouse Newsletter* (2000) reported that the passing rate on the OPT grade four reading test was 65% for Anglo-European children and 33% for African American children. Earlier grading scales left 20% of African American children passing compared to 53% of Anglo-European students, so grading was adjusted. Testing outcomes were thus an artifact of a variable performance standard and were far from absolute measurements.

A few authors compared the OPT with norm-referenced measures of school achievement but for students older than grades four and six. Robinson and Moore (1992) and Stroud (1995) found weak but significant correlations between reading and writing performance on the *Metropolitan Achievement Test* (Balow, Farr, & Hogan, 1992) and the grade nine OPT. Noel (1994) established that the grade nine OPT reading and writing tests correlated highly with the *California Achievement Tests* (CTB/McGraw-Hill, 1992) language and reading tests. Hull and Tache (1993) determined that the *Iowa Tests of Basic Skills* (Hoover, Hieronymus, Frisbie, & Dunbar, 1996) correlated significantly with OPT scores for reading and writing for 700 high schoolers.

The Research Study

Purpose

As Meisels (1989) and Popham (1999) suggested, poor scores on high-stakes testing may be due to a lack of language-based information learned in and out of school. The present study determined whether performance on a non-standardized test of curriculum mastery correlated with performance on standardized tests of oral language, writing, and reading capabilities.

Research Questions

This study answered the following questions. First, to reveal an overall impression of participants' written language and reading capabilities, did OPT reading and writing scores correlate? Second, were participants' oral and written language and reading capabilities adequate, as ascertained by standardized measures? Third, which subtests of the norm-referenced tests of oral and written language and reading correlated significantly with scores on the OPT? Fourth, which subtests of the norm-referenced tests correlated to reveal consistencies in participants' performance? Fifth, which subtests of the norm-referenced tests might predict performance on the OPT? Sixth, what language capabilities were derived as principal components of performance on the norm-referenced tests?

Method

Participants

A sample of 263 students participated in this study: 140 fourth graders (64 males and 76 females) and 123 sixth graders (56 males and 67 females). For fourth graders, the mean age was 10.1 years (range 9.3 to11.5 years) and for sixth graders the mean age was 12.1 years (range 11.2 to 13.7 years). None of the participants received special education or regular education supplemental services and parent/guardian permission was obtained for each participant.

Setting

Participants lived in an urban suburb about ten miles from the center of a moderately large city. According to the U.S. Census Bureau (2000), 90.4% of the population of the community was African American. Between one-third and one-half of African American residents were part of families that own homes. Home values averaged about $70,000 and rental properties were about $500 per month. The percent of children in this community listed as economically disadvantaged was 25.1%, compared to a state average of 13.4% (Ohio Department of Education, 2000; U.S. Census Bureau, 1990). About one-third of the children might live in homes with no husband present (U.S. Census Bureau, 1990). District enrollment was 3,100 students. Transience was noted in the schools. In 1998-1999 nearly 13% of all students were not in the district for half of the year; another 13% switched buildings in the district during the year (the state averages are about 11%) (Ohio Department of Education, 2000). Twenty-seven percent of students were eligible for free or reduced fee lunches and 31% qualified for Title I services. Special education services were provided to 30% of the student population.

The racial composition of the district staff was 41.1% non-minority, .5% Asian, and 58.4% African American (Ohio Department of Education, 1998). Although class size averaged about 24 students and annual spending per pupil was about 12% greater than the state average, the district was ranked as the fifth lowest of the county's thirty-one districts in academic accomplishments. The district met only seven of twenty-seven state academic standards. Meeting so few standards placed the district in "academic emergency," the lowest of four performance categories. About 11% of the state's districts fared this poorly.

The district met the standard for student attendance but failed the standard for graduation rate (only 75%, which was down by 5% over the past three years). In the year prior to this study, of all academic subjects tested by the OPT, the district achieved a passing rate (at least 75% of pupils tested pass) only for writing in grades six, nine, ten and twelve and reading in grades nine and ten. For grade four, 9.5% of students passed all OPT subtests; the

state average was 31.9%. For grade six, 16.1% passed all subtests; the state average was 32.5%. These were the best fourth and sixth grade scores that the district achieved over the three years prior to the study. For grade nine, 41% passed all areas (the state average was 61%), and for grade twelve 15.5% passed (state average was 39.8%). At this rate, if one were to look around an elementary classroom of twenty-four children, six will not graduate. Of the eighteen who remain, fifteen won't pass the grade twelve test. The student body can be said to have been at considerable risk for academic failure. Clearly, large numbers of students were struggling to meet state academic standards.

Instrumentation

A battery of test scores was obtained for all participants. Each test will be described.

State mandated testing. The grade four OPT reading test assessed the ability to silently read fiction and nonfiction text and (a) summarize text, (b) use graphic aids and illustrations to interpret information, (c) retell text, (d) interpret text vocabulary, (e) analyze text, (f) infer from text, (g) respond to text, (h) predict text outcomes, and (i) discern main and supporting ideas. The grade six OPT reading test included the ability to read nonfiction and fiction text and (a) analyze aspects of text, (b) summarize text, (c) infer from text, (d) respond to text, (e) critique text for its organization and logic, and (f) evaluate author's purpose.

The OPT grade four writing test required the child to prepare a letter to a friend and a personal narrative. Writers were assessed through holistic scoring for (a) clarity, (b) adherence to topic, (c) logic and organization, (d) use of a variety of words, (e) use of a variety of sentence structures, (f) awareness of word usage, and (g) correct mechanics and spelling. The grade six writing test assessed the ability to write a fictional narrative and a set of directions to complete a task. Writers were holistically scored for how well they (a) focus on a topic, (b) support ideas by giving details or examples, (c) write with logic and organization, (d) choose appropriate wording, (e) use complete and varied sentence structures, and (f) demonstrate correct mechanics and spelling. The holistic scoring rubrics yielded a score of zero to four points; only scores of three and four are passing.

Norm-referenced tests: Oral language. All oral language subtests yielded standard scores. Tests used for the grade four assessment included the *WORD-R* (Huisingh, Barrett, Zachman, Blagden, & Orman, 1990) (a test of vocabulary) and the *Test of Language Development-Intermediary* (TOLD-I:3) (Hammill & Newcomer, 1997) (verbal reasoning, grammar, listening, vocabulary). Subtests and results are found in Table 1.

Table 1. Fourth Grade Descriptive Data

Assessment	Mean Standard Score	Standard Deviation
WORD-R	100 (range 55-145)*	15*
WORD-R Synonyms	86.2	<1 SD below
WORD-R Semantic Absurdities	84.9	1 SD below
WORD-R Antonyms	88.9	<1 SD below
WORD-R Multiple Definitions	76.6	<1 SD below
TOLD-I:3	10 (range 1-20)*	3*
TOLD-I:3 Sentence Combining	10.1	Normative mean
TOLD-I:3 Picture Vocabulary	8.7	<1/2 SD below
TOLD-I:3 Generals	8.8	<1/2 SD below
TOLD-I:3 Grammatic Completion	8.8	<1/2 SD below
TOLD-I:3 Malaprops	8.4	<1 SD below
SCALE-L	50 (range 28-72)*	10*
SCALE-L All Items	44.4	1/2 SD below
TOWL-3	10 (range 1-20)*	3*
TOWL-3 Vocabulary	9.3	<1/2 SD below
TOWL-3 Sentence Combining	9.8	<1/2 SD below
TOWL-3 Story Construction	9.0	<1/2 SD below
TORC-3	10 (range 1-20)*	3*
TORC-3 General Vocabulary	8.7	<1/2 SD below
TORC-3 Syntactic Similarities	8.9	<1/2 SD below
TORC-3 Paragraph Reading	8.3	1/2 SD below

**based on normative values*

Both tests were developed using a normative sample whose demographics mirrored the United States. African Americans represent just over 12% of the population (U.S. Census Bureau, 2000). The *WORD-R* was normed using a sample that was 13% African American. For the TOLD-I:3 the normative sample was 15% African American.

Assessment of oral language in grade six included the *WORD-Adolescent Test* (WORD-A) (Zachman, Huisingh, Barrett, Orman, & Blagden, 1989) (vocabulary and semantics). The test was normed on a sample that was 14% African American. *The Test of Adolescent and Adult Language* (TOAL-3) (Hammill, Brown, Larsen, & Wiederholt, 1994) Listening Vocabulary subtest was also used. The TOAL-3 normative sample was 15% African American.

Written language. Written language measures yielded standard scores. The *Scaled Curriculum Achievement Levels Test* (SCALE) (Doherty & Roid, 1992) estimates an academic functional level for students in grades three to eight in language and reading. The test manual does not indicate the demographics for the normative sample.

The Language Usage subtest, known as the SCALE-L, assesses five curriculum strands, Composition Structure, Composition Process, Grammar/

Usage, Punctuation, and Capitalization. All items on the SCALE-L are multiple choice. Items require students to recognize written language conventions or interpret the meaning of a written message.

The second instrument used to evaluate written language was the *Test of Written Language* (TOWL-3) (Hammill & Larsen, 1996). The TOWL-3 was normed on a sample which was 13% African American.

Sixth graders' writing capabilities were further assessed via administration of two TOAL-3 subtests. Writing Vocabulary presents the same task as the TOWL-3 Vocabulary subtest (write a given word in a meaningful sentence). Writing Grammar is a test of sentence combining. The TOAL-3 was used instead of the TOWL-3 because its administration required less time—if students begin with items selected for children near their own age, not eight-year-olds, they are likely to reach the ceiling in less time.

Fourth graders were given the *Test of Reading Comprehension* (TORC-3) (Brown, Hammill, & Wiederholt, 1995). This instrument identifies the reader's ability to apply word-level vocabulary processing skills, sentence-level syntactic processing skills, and paragraph-level comprehension skills. The TORC-3 yields standard scores. The test was normed on a sample that was 12% African American.

Sixth graders' reading competencies were assessed using the SCALE Reading subtest (SCALE-R) and the reading portions of the TOAL-3. The SCALE-R assesses reading for Word Meaning, Literal Comprehension, Interpretive Comprehension, and Evaluative Comprehension. TOAL-3 Reading Vocabulary assesses single-word comprehension. Reading Grammar assesses sentence-level comprehension. The TOAL-3 was used instead of the TORC-3 because its administration required less time—if students begin reading items selected for age peers, not seven-year-old readers, they are likely to reach the ceiling in less time.

Procedures

Data collection. Standardized testing began in February and continued through early June. Each participant left class for a total of three to four hours to complete testing. Children attended two to four testing sessions in order to prevent fatigue. Testing was conducted in empty rooms in the schools. In March, students took the OPT with their classes, which was scored by the state's outsourcing arrangements and reported to the district in mid-June.

The researcher and thirty graduate students served as testers. All students had completed coursework on testing procedures, had experience administering standardized tests, and had taken a course on language development in African American children. Testers participated for fifteen to sixty-five hours. Testing teams (usually four students plus the researcher) visited the schools weekly for up to three days per week.

Groups of about ten children rotated among the testing stations. Some began with one-on-one oral language testing, while small groups of others were given the reading and written language instruments first. Tests across and within each language modality were administered in randomized order to prevent order effects. Short breaks were given as needed.

Each participant completed all measures. Tests were scored by six of the testers who served as student workers. Occasional errors made it necessary to discard some scores on some subtests. There was never more than one subtest discarded per student. All protocols were reviewed by the researcher then scores were entered into the SPSS Base 10.0 (1999) program by a graduate student trained in statistics. The researcher reviewed the SPSS files and corrected data discrepancies before final data analyses were run.

Results

Findings provided response to the research questions.

Question 1—To reveal an overall impression of the children's written language and reading capabilities, did OPT reading and writing scores correlate?

A score of 75% is required to pass any of the OPTs. In the fourth grade, 93 pupils (66.4%) achieved a score of 75% or more. Scores ranged from 31% to 98%, with a mean of 75% (standard deviation [SD] 14.6. Forty-four students (31.4%) passed the OPT writing test. Scores ranged from 37.5% to 87.5%, with a mean of 62% (SD .13). There were thirty-seven children (26.4%) who passed both subtests.

In the sixth grade, forty-three pupils (34.9%) achieved a score of 75% or more on the OPT reading test. Scores ranged from 16% to 92%, with a mean of 65.5% (SD 14.6). One hundred and one students (82%) passed the OPT writing test. Scores ranged from 12.5% to 87.5%, with a mean of 65% (SD .17). There were sixty-one children (49.5%) who passed both subtests.

Fourth grade OPT reading and OPT writing scores correlated ($r = .447$, $p \leq .0001$), as did sixth grade OPT reading and OPT writing scores ($r = .380$, $p \leq .0001$), to reveal a moderately stable and representative measure of reading and writing performance. The overall result was that students in this sample did not perform very successfully on state-mandated tests.

Question 2—Were participants' oral and written language and reading capabilities age appropriate as ascertained by standardized measures?

Scores for grade four are found in Table 1. Fourth grade mean oral language scores were consistently below normative means. Tasks that required more complex verbal reasoning and vocabulary usage posed difficulty for students. Better scores were obtained for simpler verbal tasks, such as sentence combining, picture identification, and recognizing grammatic constructions. Moderately difficult tasks, such as identifying misuse of words and

supplying synonyms and antonyms, brought scores below normative means. Deciphering multiple word meanings and semantic absurdities proved the most taxing.

Fourth grade written language was commensurate with oral language capabilities. Written sentence combining was near the normative mean, as was oral sentence combining. Vocabulary, recognizing written language meanings and conventions, and story writing were below expectations. Fairly good syntactic skills were displayed in reading for sentence meaning. However, the more complex task of reading paragraphs was an area of difficulty.

Scores for grade six are reported in Table 2. Mean scores were below expectations by varying margins. WORD-A assessed concepts related to daily living. Mean scores were close to normative means. But mean performance on TOAL-3 Listening Vocabulary, a test of multiple meaning words related to more complex concepts, was two SDs below the normative mean. This finding may signal considerable insufficiencies in vocabulary.

Table 2. Sixth Grade Descriptive Data

Assessment	Mean Standard Score	Standard Deviation
WORD-A	100 (range 55-145)*	15*
WORD-A Brand Names	95.9	<1/2 SD below
WORD-A Synonyms	91.3	1/2 SD below
WORD-A Signs of the Times	99.0	near normative mean
WORD-A Definitions	92.3	1/2 SD below
TOAL-3	10 (range 1-20)*	3*
TOAL-3 Listening Vocabulary	5.1	>1-1/2 SD below
TOAL-3 Writing Vocabulary	4.9	>1-1/2 SD below
TOAL-3 Writing Grammar	5.2	>1-1/2 SD below
TOAL-3 Reading Vocabulary	5.9	<1-1/2 SD below
TOAL-3 Reading Grammar	5.2	>1-1/2 SD below
SCALE-L	50 (range 28-72)*	10*
SCALE-L All Items	45.6	1/2 SD below
TOWL-3	10 (range 1-20)*	3*
TOWL-3 Story Construction	8.1	>1/2 SD below
SCALE-R	50 (range 28-72)*	10*
SCALE-R Level 4 Items	43.5	>1/2 SD below

**based on normative values*

Sixth graders showed notable insufficiencies in written language. TOAL-3 scores indicated that vocabulary usage and sentence construction were considerably below norms. The SCALE-L showed that participants' knowledge of written language conventions was not strong. The SCALE-R showed difficulties in reading for word meaning and passage comprehension.

Question 3—Which subtests of the norm-referenced tests of oral and written language and reading correlated significantly with scores on the OPT?

Subtest means that correlated at $p \leq .0001$, $r \geq .4$ (Pearson Product Moment correlations) are found in Tables 3 and 4. For grade four, several oral language mean scores correlated with OPT reading scores. Oral language development appeared relevant to test performance. Written language means also correlated with OPT reading scores. Knowledge of written language conventions, vocabulary, and sentence manipulation skills appeared related to performance on the mandated reading test. Correlations between OPT reading and standardized reading test scores occurred. Notably, a reading score, TORC-3 General Vocabulary, was the only standardized measure that correlated with the OPT writing score.

For sixth grade, oral and written language scores correlated with OPT reading, again attesting to the interrelatedness of these various language skills. There were no grade six reading measures that correlated with OPT reading. None of the grade six measures correlated with OPT writing. This points to the possible spuriousness of the OPT writing scores in comparison to the other tests given. It appears that the OPT writing mean was anomalously high.

Analysis of these correlations suggested that, in both grades, (a) vocabulary-driven measures of oral and written language correlated frequently and strongly with the OPT tests, and (b) subtests that measured written language usage and conventions correlated with the state's measure of reading abilities.

Question 4—Which subtests of the norm-referenced tests correlated to reveal consistencies in participants' performance?

Subtest means that correlated at $p \leq .0001$, $r \geq .4$ are found in Tables 3 and 4. For fourth graders, numerous correlations within tests affirmed the strength of the instruments. Correspondences between vocabulary and verbal reasoning were seen within and across oral and written language modalities. Reading scores correlated within the reading instrument and with oral and written language testing.

Correlations between vocabulary-driven oral language measures were found. Several measures of oral vocabulary and linguistic reasoning correlated with written vocabulary and written language scores. Correlations were evidenced between two sorts of composition skills: conceptual (meaning-based) and mechanical. Reading vocabulary correlated with oral vocabulary, oral measures of linguistic reasoning, and with overall written language skills. Reading comprehension at the sentence level correlated with oral tasks involving linguistic reasoning and with written sentence combining. Sentence processing and sentence construction were related skills.

For grade six, numerous oral language subtests correlated with one another, as did written language scores. There were multiple correlations

Table 3. Fourth Grade Correlation Matrix

	OPT RD	OPT WR	WORD SYN	WORD SEM	WORD ANT	WORD MDEF	TOLD SENCO	TOLD PVOC	TOLD GEN	TOLD GRM	TOLD MALP	SCALE L	TOWL VOCAB	TOWL SENCO	TOWL STORY	TOR C VOC	TORC SYNTX	TORC PARA
OPTRD			.594	.537	.438	.550			.530		.434	.546	.485	.406		.451	.406	.422
OPT WRIT																.444		
WORD SYN	.594			.649	.674	.515		.532	.546		.489	.545	.526			.515		
WORD SEM	.537		.649		.677	.509	.400	.500	.605		.441	.426	.453			.431	.401	
WORD ANT	.438		.674	.677		.466		.441	.517		.449		.453			.449	.480	
WORD MDEF	.550		.515	.509	.466			.429		.409	.467	.624	.494			.459		
TOLD SENCO			.400															
TOLD PVOC			.532	.500	.441	.429			.477		.496					.400		
TOLD GEN	.530		.546	.605	.517			.477			.443		.437			.496		
TOLD GRM						.409						.403						
TOLD MALP	.434		.489	.441	.449	.467		.496	.443				.520			.413		
SCALE	.546		.545	.426		.624				.403			.501			.454		
TOWL VOCAB	.485		.526	.453	.453	.494			.437		.520	.501				.543		
TOWL SENCO	.406																.414	.457
TOWL STORY																		
TORC VOC	.451	.444	.515	.431	.449	.459		.400	.496		.413	.454	.543	.414				.458
TORC SYNTX	.406		.401	.480										.457		.458		
TORC PARA	.422																	

Note. Pearson r, significant at p≤ .0001

Table 4. Sixth Grade Correlation Matrix

	OPT RD	OPT WR	WORD BRND	WORD SYN	WORD SOT	WORD DEF	TOAL LISTN	SCALE L	TOWL STORY	TOAL WVOC	TOAL WGRM	TOAL RVOCB	TOAL RGRM	SCALER
OPT READ				.458		.500	.418	.590		.455				
OPT WRIT														
WORD BRND				.571	.575	.593	.413							.444
WORD SYN	.458		.571			.773	.485	.466		.601		.548		.464
WORD SOT			.575			.561	.453							
WORD DEF	.500		.593	.773	.561		.498	.474		.552		.535		.499
TOAL LISTN	.418		.413	.485	.453	.498		.492		.555		.498		
SCALE L	.590			.466		.474	.492			.599	.465	.569	.484	
TOWL STORY														
TOAL WVOC	.455			.601		.552	.555	.599			.447	.568		
TOAL WGRM								.465		.447				
TOAL RVOCB				.548		.535	.498	.569		.568			.414	.569
TOAL RGRM								.484				.414		.484
SCALE R			.444	.464		.499						.569	.484	

Note. Pearson *r*, significant at *p*≤ .0001

between vocabulary-driven oral measures. Vocabulary correlated with measures of conceptual understanding. Correlations between conceptual and mechanical composition skills were evidenced. Sentence combining skill correlated with other measures of written language skill.

Multiple measures of reading comprehension correlated, demonstrating the relationship between comprehension at the word, sentence, and paragraph levels. Reading vocabulary was related to oral vocabulary and to oral linguistic reasoning. Recognition of written language conventions and usage correlated with reading comprehension at the sentence level.

Question 5—Which subtests of the norm-referenced tests might predict performance on the OPT?

For grade four, linear regressions were run for the measures that correlated significantly ($p \leq .0001$) with the OPT reading and OPT writing scores. Predictors of OPT reading performance were WORD-R Synonyms ($R^2 = .353$), WORD-R Multiple Definitions ($R^2 = .303$), WORD-R Semantic Absurdities ($R^2 = .288$), TOLD-I:3 Generals ($R^2 = .281$), TOWL-3 Vocabulary ($R^2 = .235$), TORC-3 General Vocabulary ($R^2 = .204$), WORD-R Antonyms ($R^2 = .191$), TOLD-I:3 Malaprops ($R^2 = .189$), TORC-3 Paragraph Reading ($R^2 = .178$), TOWL-3 Sentence Combining ($R^2 = .165$), and TORC-3 Syntactic Similarities ($R^2 = .164$). Only TORC-3 General Vocabulary ($R^2 = .197$) was predictive of the OPT writing outcome.

To determine whether several factors taken together accounted for a percent of the variance in OPT reading scores for fourth graders, multiple regressions were run. Combinations of the variables that yielded the most powerful linear regressions were applied. To account for the impact of oral vocabulary as a variable in OPT reading performance, a multiple regression was run using the following scores as independent variables: WORD-R Synonyms, WORD-R Semantic Absurdities, WORD-R Antonyms, WORD-R Multiple Definitions, TOLD-I:3 Generals, and TOLD-I:3 Malaprops. The R^2 value for these factors was .458, $p \leq .0001$. When taken together, these vocabulary capabilities accounted for roughly 46% of the variance in OPT reading scores. Another multiple regression was run using TORC-3 General Vocabulary and TOWL-3 Vocabulary as independent variables. The R^2 value for these factors was .285, $p \leq .0001$, suggesting that reading and writing vocabulary accounted for about 29% of the variance in OPT reading performance. An additional multiple regression used TORC-3 Syntactic Similarities and TORC-3 Paragraph Reading as independent variables. The R^2 value for measures that required complex reading was .268, $p \leq .0001$, accounting for about 27% of the variance in OPT reading scores. Finally, a multiple regression was run using TOWL-3 Vocabulary and TOWL-3 Sentence Construction as independent variables, to explore the impact of written response quality on OPT reading outcomes. The R^2 value was .301, $p \leq .0001$, which accounted for about 30%

of the variance in OPT reading performance. Reading ability itself was not more important than vocabulary and the ability to use words and sentences to construct written responses.

Regarding sixth grade data, linear regressions were run for the measures that correlated ($p \leq .0001$) with OPT reading at $r = .4$ or better. Predictors were SCALE-L ($R^2 = .348$), WORD-A Definitions ($R^2 = .250$), WORD-A Synonyms ($R^2 = .210$), TOAL-3 Writing Vocabulary ($R^2 = .207$), and TOAL-3 Listening Vocabulary($R^2 = .175$).

Multiple regressions determined whether several factors taken together accounted for a percent of the variance in OPT reading scores for sixth graders. One multiple regression was run using TOAL-3 Listening Vocabulary and TOAL-3 Writing Vocabulary as independent variables. The R^2 value for these factors was .244, $p \leq .0001$, denoting that complex oral and written vocabulary knowledge and usage accounted for 25% of the variance in OPT reading scores. Another multiple regression used only oral vocabulary measures, TOAL-3 Listening Vocabulary, WORD-A Synonyms, WORD-A Definitions. The R^2 value for these factors was .292, $p \leq .0001$, accounting for approximately 30% of the variance in OPT reading outcomes. An additional multiple regression used the SCALE-L and SCALE-R as independent variables, to explore the importance of knowledge of written language conventions and reading abilities. The R^2 value for these variables was .347, $p \leq .0001$, suggesting that 35% of the variance in performance could be attributed to reading and written language abilities. A final regression used the TOAL-3 Writing Vocabulary and the SCALE-L as independent variables. The R^2 value for these variables was .363, $p \leq .0001$, denoting that written language capabilites accounted for over 36% of the variance in OPT reading performance. For sixth graders, comprehension of the meaning of written words as well as knowledge of the mechanics and conventions of written language were important factors, slightly overshadowing oral vocabulary alone as predictive variables.

For both grades, OPT reading performance can be predicted by word knowledge and word usage skills. The ability to manipulate compositional elements (i.e., choosing and ordering words to form sentences, using sentences to express multi-sentence ideas, choosing how to best convey a purpose, writing mechanics) had some predictive power. Reading comprehension testing was also predictive.

Question 6—What language capabilities were derived as principal components of performance on the norm-referenced tests?

Principal Components Analysis (PCA) was performed to determine the number and composition of language capabilities that were operative as children took the norm-referenced tests. PCA reduces the number of variables analyzed and identifies basic components of performance. This statistic determines the fundamental dimensions tapped by larger sets of variables, in this

case the multiple normed subtests. PCA was performed separately by grade.

The number of components was determined using the Eigenvalue greater than 1 criteria as well as the logical interpretability of the resulting components (Tabachnick & Fidell, 1996). In fourth grade three factors representing language abilities accounted for 53% of the variance in scores. The first factor was labeled Vocabulary and had high loadings from WORD-R Synonyms, WORD-R Semantic Absurdities, WORD-R Antonyms, WORD-R Multiple Definitions, TORC-3 General Vocabulary, TOWL-3 Vocabulary, TOWL-3 Story Construction, TOLD-I:3 Sentence Combining, TOLD-I:3 Picture Vocabulary, TOLD-I:3 Generals, TOLD-I:3 Grammatic Completion, and TOLD-I:3 Malaprops. The second factor was labeled Written Language and had high loadings from SCALE-L. The third factor was labeled Syntax and had high loadings from TORC-3 Syntactic Similarities and TOWL-3 Sentence Combining.

For sixth grade three factors accounted for 53% of the variance in scores. The first factor was labeled Written Language and consisted of high loadings from SCALE-L, TOAL-3 Reading Vocabulary, TOAL-3 Reading Grammar, TOAL-3 Writing Vocabulary, and TOAL-3 Writing Grammar. The second factor was labeled Vocabulary and consisted of high loadings from WORD-A Brand Names, WORD-A Synonyms, WORD-A Signs of the Times, WORD-A Definitions, and TOAL-3 Listening Vocabulary. This factor also had a high cross loading from TOAL-3 Writing Vocabulary. The third factor was labeled Reading Comprehension and consisted of high loadings from SCALE-R. This factor also had a substantial cross loading from SCALE-L.

Factor scores were computed for each component and these scores served as composite variables in subsequent analyses. Factor scores derived from the above PCAs were entered simultaneously as independent variables in regression analyses, for the purpose of predicting OPT reading and OPT writing scores. In fourth grade, all language ability factors were significant predictors of OPT reading scores (Vocabulary _=.42, $t(125)=5.42$, $p\leq.001$; Written Language _=.30, $t(125)=3.88$, $p\leq.001$; _=.18, Syntax $t(125)=2.61$, $p=.01$). Together these variables accounted for 46% of the variance in OPT reading scores (total Regression $F(3,125)=35.27$, $p\leq.001$). Only Vocabulary and Written Language predicted unique variance in OPT writing scores (Vocabulary _=.21, $t(125)=2.21$, $p=.03$; Written Language _=.22, $t(125)=2.29$, $p=.02$; Syntax _=.07, $t(125)=0.82$, $p=.41$). Together these variables accounted for 15% of the variance in OPT writing scores (total Regression $F(3,125)=7.54$, $p\leq.001$). Language abilities accounted for substantially more variance in OPT reading than in OPT writing scores ($t(126)=4.05$, $p\leq.001$).

In grade six, all language ability factors predicted significant unique variance in OPT reading scores (Vocabulary _=.22, $t(110)=2.69$, $p=.008$; Written Language _=.37, $t(110)=4.47$, $p\leq.001$; _=.24, Reading Comprehension $t(110)=2.82$, $p=.006$). Together these variables accounted for 44% of the vari-

ance in OPT reading scores (total Regression $F(3,110)=28.30$, $p\leq.001$). Only Vocabulary abilities were a marginally significant predictor of unique variance in OPT writing scores (Vocabulary _=.19, $t(110)=1.81$, $p=.07$; Written Language _=.12, $t(110)=1.14$, $p=.26$; Reading Comprehension _=.10, $t(110)=0.92$, $p=.36$). However, together these variables accounted for 11% of the variance in OPT writing scores (total Regression $F(3,110)=4.33$, $p=.006$). Again, language abilities accounted for substantially more variance in OPT reading scores than in OPT writing scores ($t(111)=4.12$, $p\leq.001$).

Discussion

Limitations of the Study

This study was limited to African American students in one community. This participant selection factor may limit generalizability of findings to other groups and other communities. Results are pertinent to the OPT and cannot be applied to other state, commercial, or NAEP assessments.

Implications

Findings yield several implications for preparing students for mandated testing and for identifying and serving students who may be at risk for less adequate performance. First, the importance of building students' vocabularies cannot be underestimated. Results demonstrated the importance of vocabulary knowledge to reading comprehension (Johnson & Pearson, 1978; Nagy, 1988). Second, reading scores coincided with the ability to manipulate compositional elements. Knowledge of written language mechanics and conventions consistently correlated with reading scores. Students must learn how the mechanical conventions of language are used to convey meaning.

Third, oral and written language and reading scores were low, placing students at considerable risk for poor performance on high-stakes testing. Other studies have shown concerns relative to test performance for minority and low income students, especially when tests scrutinized verbal abilities (Morris, 2000; NAEP, 1998; NAEP, 1999; NAEP, 2000; Schellenberg, 1998; Winfield, 1990). The present study is significant in its relationship to published documentation which proposed that risk of academic failure can be linked to residential and school segregation (Jackson, 1999). Students sampled may have had a somewhat more elevated risk for school achievement concerns.

Fourth, the interrelationship of language development and academic achievement is apparent. Speech-language pathologists, reading specialists, learning disabilities specialists, and special educators can be called upon to assist regular education teachers in earlier identification of academically-relevant language deficiencies several years before children undergo grade four achievement testing.

Conclusions

The interrelationship of measures of language development and academic achievement has been demonstrated. Consistencies in performance were related to the underlying element of vocabulary. Future research on the relationship between language development and academic achievement is needed.

The USDOE (2001) suggests that innovations in high-stakes testing should include developing assessments that can yield better information about areas of school achievement that students may be struggling with and that can diagnose learners' specific problems. Perhaps this will result in the creation of instruments that are sensitive to how language learning impacts academic achievement.

References

Balow, I. H., Farr, R. C., & Hogan, T. P. (1992). *Metropolitan Achievement Test.* San Antonio, TX: Harcourt Brace Educational Measurement.

Barkley, R. (2001). The OEA [Ohio Education Association] on testing. *Ohio Schools, 79,* 12.

Bobbett, G. C. (1993, April). *The impact of community/school characteristics on student outcomes: An analysis of report cards on schools.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Brown, V. L., Hammill, D. D., & Wiederholt, J. L. (1995). *Test of Reading Comprehension (TORC-3).* Austin, TX: Pro-ed.

Butler, K. (1999). From oracy to literacy: Changing clinical perceptions. *Topics in Language Disorders, 20*(1), 14-32.

Commission on Behavioral and Social Sciences and Education. (1998). *Equivalency and Linkage of Educational Tests Interim Report.* Washington, DC: National Academy Press. Retrieved August 5, 2001, from http://books.nap.edu/books/N1000906html/32.html

CTB/McGraw-Hill. (1992). *California Achievement Test.* Monterey, CA: CTB/McGraw-Hill.

Denoyer, R., & White, M. (1992). *Proficiency testing and the income gap.* (ERIC Document Reproduction Service No. ED35064)

Doherty, V. W., & Roid, G. H. (1992). *Scaled Curriculum Achievement Levels Test (SCALE).* Los Angeles, CA: Western Psychological Services.

Falk-Ross, F. C. (2002). *Classroom-based language and literacy intervention: A program and case study approach.* Boston: Allyn & Bacon/Longman.

Gallagher, M. P. (1993, April). *Proficiency testing and poverty: Looking within a large urban district.* Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA.

Glaser, R. (1994). Criterion-referenced tests: Part II. Unfinished business. *Educational Measurement: Issues and Practice, 13*(4), 27-30.

Hammill, D. D., Brown, V. L., Larsen, S. C., & Wiederholt, J. L. (1994). *Test of Adolescent and Adult Language (TOAL-3).* Austin, TX: Pro-ed.

Hammill, D. D., & Larsen, S. C. (1996). *Test of Written Language (TOWL-3)*. Austin, TX: Pro-ed.

Hammill, D. D., & Newcomer, P. L. (1997). *Test of Language Development-Intermediate (TOLD-I:3)*. Austin, TX: Pro-ed.

Hoover, H. D., Hieronymus, A. N., Frisbie, D. A., & Dunbar, S. B. (1996). *Iowa Test of Basic Skills*. Itasca, IL: Riverside Publishing.

Huisingh, R., Barrett, M., Zachman, L., Blagden, C., & Orman, J. (1990). *The WORD-R: A test of expressive vocabulary and semantics*. East Moline, IL: Linguisystems, Inc.

Hull, M., & Tache, D. (1993). *Are Iowa Test of Basic Skills stanines predictors of success on Ohio's Ninth Grade Proficiency Test?* (ERIC Document Reproduction Service No. ED35064)

International Reading Association (IRA). (1999). High stakes assessments in reading. *The Reading Teacher, 53*(3), 257-263.

Jackson, J. F. (1999). What are the real risk factors for African American children? *Phi Delta Kappan, 81*(4), 308-312.

Johnson, D., & Pearson, P. D. (1978). *Teaching reading vocabulary*. New York: Holt, Rinehart & Winston.

Kohn, A. (2000). *The case against standardized testing*. Portsmouth, NH: Heinemann.

Ladson-Billings, G. (1994). *The dreamkeepers*. San Francisco: Jossey-Bass Publishers.

Lanese, J. F. (1992, April). *Statewide proficiency testing: Establishing standards or barriers?* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Manning, M. L., Lucking, R., & MacDonald, R. H. (1995). What works in urban middle schools. *Childhood Education, 71*(4), 221-224.

McGee, G. W. (1997, March). *What state tests test*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Meisels, S. (1989). *Testing, tracking, and retaining young children: An analysis of research and social policy*. Washington, DC: National Center for Education Statistics.

Morris, D. R. (2000). *Assessing the implementation of high stakes reform: Aggregate relationships between retention rates and test results*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Nagy, W. E. (1988). *Vocabulary instruction and reading comprehension (Tech. Rep. No. 431)*. Champaign, IL: University of Illinois, Center for the Study of Reading.

National Assessment of Educational Progress (NAEP). (1998). *Long-term trends in student reading performance*. Retrieved August 5, 2001, from http://nces.ed.gov/pubsearch

National Assessment of Educational Progress (NAEP). (1999). *NAEP 1998 reading report card for the nation and the states*. Retrieved August 5, 2001, from http://nces.ed.gov/pubsearch

National Assessment of Educational Progress (NAEP). (2000). *The NAEP guide*. Retrieved August 5, 2001, from http://nces.ed.gov/pubsearch

National Center on Educational Outcomes (NCEO). (2001). *Special topic areas*. Retrieved August 5, 2001, from http://www.coled.umn.edu/nceo

National Education Goals Panel (NEGP). (1994). *Goals 2000 The National Education Goals Report: Building a nation of learners 1994*. Washington, DC: Author.

National Educational Research Policy and Priorities Board. (1999). *Investing in learn-*

ing: A policy statement on research in education. Washington, DC: US Department of Education.

Noel, A. W. (1994). *Predictors of success on Ohio's Ninth Grade Proficiency Test*. (ERIC Document Reproduction Service No. ED392803)

Ohio Department of Education. (1995). *Fourth grade proficiency tests: Information guide*. Columbus, OH: Author.

Ohio Department of Education. (1998). *Staff profile*. Retrieved August 12, 2003, from http://www.ode.state.oh.us/htbinstaff_Profile.com?SELECTION=IRN&irn=045005+Warrensville+Heights+City+SD+%28Cuyahoga%29&county=01+Adams&ctytot=01+Adams&vepd=001+Apollo+JVSD+VEPD&years=1998

Ohio Department of Education. (2000). *State of Ohio 2000 school district report card*. Retrieved August 6, 2001, from http://www.ode.state.oh.us

Ohio Statehouse Newsletter. (2000, May). Columbus, OH: Author.

Ohio Student Testing Requirements. (1991). Columbus, OH: Ohio State Legislative Office of Education Oversight.

Popham, W. (1994). The instructional consequences of criterion-referenced clarity. *Educational Measurement: Issues and Practice, 13*, 15-18, 30.

Popham, W. (1999). Why standardized tests don't measure educational quality. *Educational Leadership, 56*(6), 8-15.

Pottle, P. (2001). Coping with a process-oriented mandate. *Primary Voices K-6, 9*(3), 32-37.

Raivetz, M. J. (1992). Can school districts survive the politics of state testing initiatives. *NASSP Bulletin, 76*(545), 57-65.

Robinson, M. A., & Moore, M. H. (1992, April). *State proficiency tests and standardized tests*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Roderick, M. (1995). *Grade retention and school dropout: Policy debate and research questions*. Phi Delta Kappa Center for Evaluation, Development, and Research, Research Bulletin No. 15. Retrieved June 16, 2001, from http://www.pdkintl.org/edres/resbul15.htm

Schellenberg, S. J. (1998, April). *Does it matter where poor kids live? A look at concentrated poverty and achievement*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.

Snodgrass, D. G., & Salzman, J. A. (1998). *Graphical representations of data: Iowa Tests of Basic Skills, Ohio Sixth Grade Proficiency Tests, and SB 55*. Paper presented at the Mid-Western Educational Research Association Annual Conference, Chicago.

Snow, C. E., Burns, S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

SPSS Base 10.0 Applications Guide. (1999). Chicago: SPSS, Inc.

Stroud, R. E. (1995). *Correlations between the Metropolitan Achievement Tests, Seventh Edition, and the Ohio Ninth-Grade Proficiency Tests*. (ERIC Document Reproduction Service No. ED394992)

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: HarperCollins.

U.S. Census Bureau. (1990). *U.S. Census Data*. Database C90STF1A. Retrieved June 16, 2001, from www.venus.census.gov

U.S. Census Bureau. (2000). *American FactFinder*. Retrieved June 16, 2001, from http://factfinder.census.gov

United States Department of Education (DOE). (2001). *Back to school, moving forward*.

Retrieved December 5, 2002, from http://www.ed.gov/inits/backtoschool/index.html

United States Department of Education (DOE). (2002). *No child left behind.* Retrieved December 5, 2002, from http://www.NoChildLeftBehind.gov/

Urdan, T., & Davis, H. (1998). *Differences by race and grade level in motivation for taking standardized achievement tests.* Paper presented at the Bi-annual Meeting of the Society for Research on Adolescence, San Diego, CA.

Westbury, M. (1994). The effect of elementary grade retention on subsequent school achievement and ability. *Canadian Journal of Education, 19*(3), 241-250.

Winfield, L. F. (1990). School competency testing reforms and student achievement: Exploring a national perspective. *Educational Evaluation and Policy Analysis, 12*(2), 157-173.

Zachman, L., Huisingh, R., Barrett, M., Orman, J., & Blagden, C. (1989). *The WORD-Adolescent Test (WORD-A): A test of expressive vocabulary and semantics.* East Moline, IL: Linguisystems, Inc.

Acknowledgments

This research was supported by a New Faculty Research Award, NF-001, Cleveland State University.

The author thanks Anna Lisa Lincoln, statistician.