TEMI K-2

Texas Early Mathematics Inventories

Technical Manual

for Progress Monitoring and Outcome Tests

© 2008 University of Texas System/Texas Education Agency

Technical Manual
for the
Texas Early Mathematics Inventories – Progress Monitoring
and
Texas Early Mathematics Inventories – Outcome
Grades K, 1, and 2

Brian R. Bryant
Diane Pedrotty Bryant
Minyi Shih
Rick Sperling
Sun a Kim
You-Jin Seo

© 2008 The University of Texas System/Texas Education Agency

This manual provides information regarding the technical properties of the *Texas Early Mathematics Inventories – Progress Monitoring* (TEMI-PM) and the *Texas Early Mathematics Inventories – Outcome* (TEMI-O). In this manual, we provide information pertaining to (a) the norms, (b) reliability, and (c) validity of the test scores.

Norms

In this section, we provide information about how the tests were normed. Specifically, we discuss (a) the procedures we used to select participants, (b) the demographic characteristics of the sample, (c) the methods we used to equate the alternate forms of the measures, and (d) the types of normative scores available on the TEMI-PM and the TEMI-O.

Selection Procedures

In the spring of 2007, we devised a plan to solicit schools to participate in the normative process. First, we participated in a Texas Education Telecommunications Network (TETN) webinar, during which we presented an overview of the 3-Tier Mathematics Project, introduced the tests, told of an upcoming solicitation webcast, and asked education service center (ESC) personnel to help us identify schools in their regions that might be willing to participate in the norming process.

We also enlisted the assistance of the Texas Elementary Principals and Supervisors Association. We described our project and asked that information about the project and the webcast be posted on the organization's Web site and be advertised in its newsletter.

In May 2007, we presented a 45-minute webcast describing the project, the tests, and the recruitment drive. We would provide tests for one classroom each in grades K, 1, and 2; classroom teachers would administer the tests after watching a training CD; and test protocols would be returned to us. We would score the tests and provide written reports back to each school within 3 weeks.

During and after the webcast, we received numerous questions via e-mail, and answered them all in a timely fashion. As part of the webcast, we provided our Web site's URL, where schools could go for more information and also to register to participate in the testing, which would be conducted in grades K, 1, and 2, in September and October 2007, January 2008, and May 2008. Our goal was to have at least 1,500 students represented in each grade and have all 20 ESCs represented.

Within 6 weeks, we had more than 80 schools registered and 18 of the 20 ESCs represented. In August 2007, we contacted colleagues from the two ESCs where we still lacked participants, and we were given the names of principals at schools in the regions who might be interested. After contacting the principals and describing the project, we received commitments from all 20 ESCs.

Demographic Characteristics

The TEMI-PM and TEMI-O were co-normed on students across all 20 ESCs in Texas (kindergarten = 1,791 students, grade 1 = 1,764 students, and grade 2 = 1,675 students; see Figure 1). The following school districts contributed to the development and standardization process: Agua Dulce Independent School District (ISD), Angleton ISD, Arlington ISD, Athens ISD, Birdville ISD, Bosqueville ISD, Bovina ISD, Brownfield ISD, Calhoun County ISD, Central ISD, Cisco ISD, Clarksville ISD, Como-Pickton Consolidated Independent School

© 2008 The University of Texas System/Texas Education Agency

District (CISD), Comstock ISD, Conroe ISD, Copperas Cove ISD, Corsicana ISD, Crockett County Consolidated Common School District (CCSD), Culberson County-Allamoore ISD, Del Valle ISD, Devine ISD, East Central ISD, Flour Bluff ISD, Floydada ISD, Granbury ISD, Grapeland ISD, Hamilton ISD, Hardin-Jefferson ISD, Howe ISD, Hubbard ISD, Hughes Springs ISD, Irion County ISD, Jefferson ISD, Joaquin ISD, Kaufman ISD, Kingsville ISD, Kopperl ISD, Lackland ISD, Lamesa ISD, Levelland ISD, Liberty-Eylau ISD, Lovelady ISD, Lubbock-Cooper ISD, Lufkin ISD, Marfa ISD, Mason ISD, McDade ISD, McGregor ISD, Mildred ISD, Moran ISD, New Braunfels ISD, North Forest ISD, North Hopkins ISD, Olney ISD, Panther Creek CISD, Paris ISD, Pflugerville ISD, Pittsburg ISD, Poteet ISD, Prosper ISD, Quinlan ISD, Robinson ISD, San Benito CISD, San Elizario ISD, Shepherd ISD, Socorro ISD, Stephenville ISD, Stockdale ISD, Tarkington ISD, Tenaha ISD, Valley View ISD, Weatherford ISD, Winfield ISD, and Ysleta ISD.

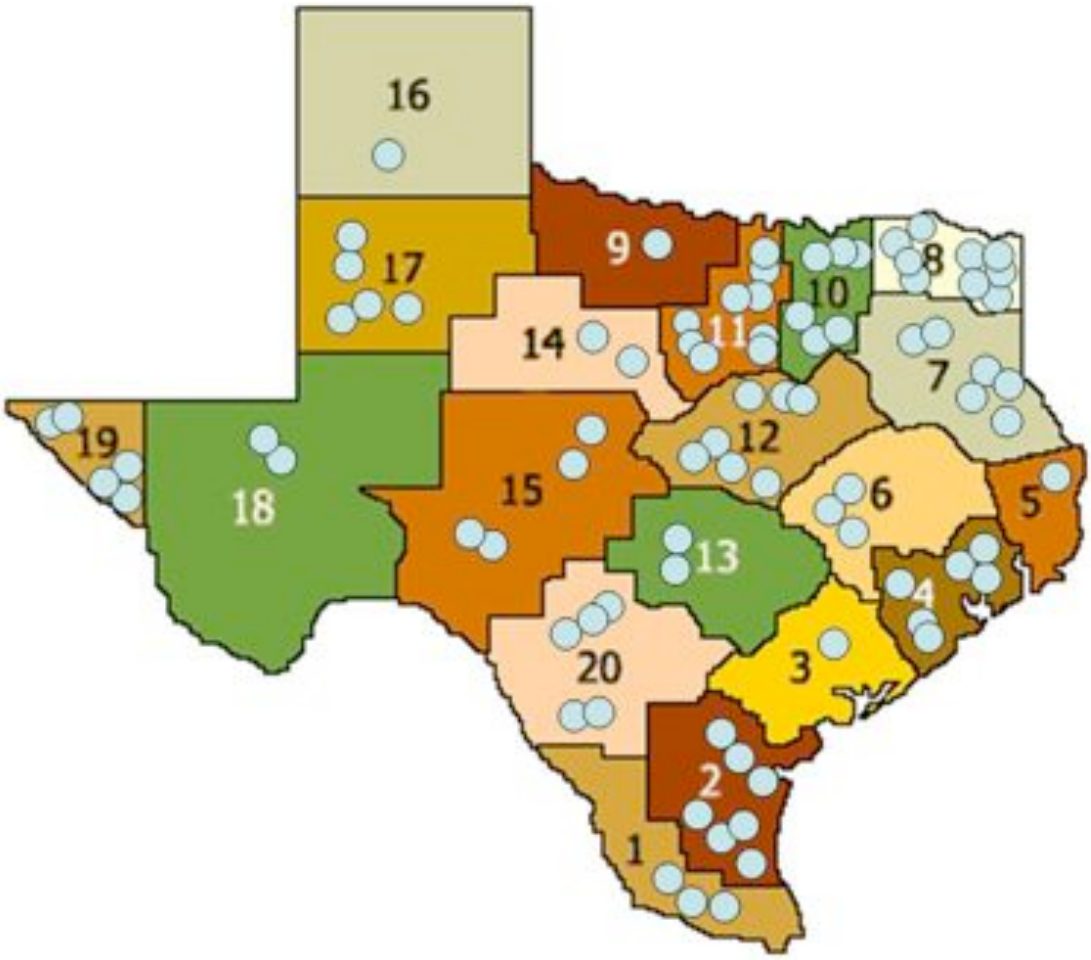

Figure 1. Standardization Sites Across Texas

The demographic characteristics of the normative sample, by grade, are found in Tables 1, 2, and 3. The demographics were adjusted to account for the percentage of Spanish-speaking students attending classes for English language learners (ELLs). The TEMI-PM and the TEMI-O are intended for English-speaking students only. At the time of the writing of this manual, a

© 2008 The University of Texas System/Texas Education Agency

Spanish version of the test was being field tested. Thus, based on information for the Texas Education Agency (TEA), we adjusted the numbers in each ethnic category to account for only the students for whom the tests are intended, and those figures are reported in the demographic tables.

Table 1. Ethnic and Gender Demographic Characteristics for Kindergarten Sample

	Grade K State Percentage	Grade K Adjusted State Percentage	Total	Final Percentage
African American	13.20%	13.80%	247	13.80%
Asian	3.40%	3.50%	63	3.50%
Native American	0.40%	0.45%	8	0.45%
Hispanic	50.20%	45.85%	821	45.85%
White	32.80%	36.40%	652	36.40%
Male	51.7%		926	51.70%
Female	48.3%		865	48.30%

Table 2. Ethnic and Gender Demographic Characteristics for Grade 1 Sample

	Grade 1 State Percentage	Grade 1 Adjusted State Percentage	Total	Final Percentage
African American	13.60%	14.60%	258	14.60%
Asian	3.50%	3.77%	67	3.77%
Native American	0.40%	0.48%	8	0.48%
Hispanic	50.30%	46.80%	826	46.80%
White	32.40%	34.35%	606	34.35%
Male	51.8%		914	51.81%
Female	48.2%		850	48.19%

Table 3. Ethnic and Gender Demographic Characteristics for Grade 2 Sample

	Grade 2 State Percentage	Grade 2 Adjusted State Percentage	Total	Final Percentage
African American	13.80%	14.80%	248	14.80%
Asian	3.50%	3.74%	63	3.74%
Native American	0.30%	0.34%	6	0.34%
Hispanic	49.20%	45.78%	767	45.78%
White	33.20%	35.34%	592	35.34%
Male	51.3%		859	51.28%
Female	48.7%		816	48.72%

© 2008 The University of Texas System/Texas Education Agency

Equating Procedures

First, in the summer of 2005 we created Form A for all measures, TEMI-PM and TEMI-O. For the TEMI-PM, Form B was created by making slight changes to Form A. For example, if a Magnitude Comparisons (MC) item contained the numerals 3 and 0, the corresponding item on Form B contained the numerals 3 and 1. This procedure was used for all TEMI-PM subtests. We then created Forms C and D in the same way, by examining the two existing forms and making minor alterations to create items of similar difficulty (e.g., 2 and 0, and 4 and 0).

For the TEMI-O, Form A was created for the Mathematics Problem Solving (MPS) and Computation subtests. For MPS, Form B was created, employing items similar to Form A, with minor adjustments made similar to with the TEMI-PM. For example, if the Form A test item called for a student to count six cookies and mark the numeral that goes with the total number of cookies, the corresponding Form B item had the student count seven cookies. The remaining items were generated in similar fashion. Occasionally, we created a new item that assessed the same construct on Form B in a slightly different way than was done on Form A, but this was rarely done. For Computation, two items of a similar nature were placed side by side (5 + 3 for Form A became 6 + 3 for Form B). Later, Form C of each subtest was created using the same procedure.

The multiple forms were administered in the 2005–2006 school year. Item difficulty levels were used to equate the forms. On the TEMI-PM, after each set of five items, we compared mean scores to ensure they were similar across all forms. When one or another form leaped ahead or lagged behind, we exchanged items across forms to balance out the scores. For example, we might have taken item 6 from Form D and exchanged it with item 3 from Form B, based on their percentages of difficulty. We recomputed the averages across each five-item set, and continued the process until the mean scores were similar. We did this with every five-item set across all forms, and ended up with four forms that had equivalent means and standard deviations.

For the MPS subtest of the TEMI-O, we did something a bit different than what we did on the TEMI-PM. Instead of moving items from one form to the other, we changed the item slightly to make it either easier or more difficult. For example, if for some reason item 6 on Form A were easier than the corresponding item on Form B, we made the Form B item easier (e.g., by changing the number of cookies from seven to five). For Computation, the subtest means were virtually identical, so we left the forms as they were; there was no need to make changes.

We then used the final versions of the test in the winter and spring of 2006, with one exception. In the spring, we changed to a group format for the TEMI-PM. This required us to re-examine the item scores across forms to ensure that they were still equivalent, which they were. We thus were able to go into the 2006–2007 school year to field test the experimental versions further. In that year, we administered all four forms of the TEMI-PM and TEMI-O.

TEMI-PM and TEMI-O Scores

Two types of scores are used with the TEMI-PM and the TEMI-O: Student Scores and Percentiles. Each is described here.

Student Scores

For the TEMI-PM and the TEMI-O, Student Scores—or raw scores, as they are often called—represent the total number of points attained on a subtest (e.g., Magnitude Comparisons, Number Sequences) or the summed values of all the subscales (i.e., the Total Score of the TEMI-

© 2008 The University of Texas System/Texas Education Agency

PM or TEMI-O. These scores are used for progress-monitoring purposes and to derive Percentile Ranks.

Percentiles

Student Scores are converted to Percentiles using the Normative Tables found in the Teacher's Manual. Percentiles—or Percentile Ranks, as they are also called—range from 1 through 99 and represent the percentage of students in the normative sample who score at or below the Student Score. For example, for Magnitude Comparisons in grade 1, a percentile of 35 means that 35% of the TEMI-PM normative sample scored at or below that Student Score. Percentiles are popular scores because they are easy to interpret and understand. Percentiles should not be confused with percentage correct. A Percentile Rank of 43 means that 42% of the normative sample scored at or below the associated student score. It does not mean that a student answered 42% of the items correctly. The table below refers to descriptive ratings that coincide with percentile rank ranges.

Table 4. Example of Relationship Between Percentile Ranks and Descriptive Ratings

%ile Rank Range	Descriptive Rating	% in Normative Sample
< 2	Very Poor	1.99
2–8	Poor	7
9–24	Below Average	16
25–75	Average	50
76–86	Above Average	16
87–98	Superior	7
> 98	Very Superior	1.99

Of particular interest with the TEMI-PM and the TEMI-O is the 25th percentile, which represents the lower end of Average performance. Below that point, if the teacher verifies the performance level, the student is eligible for intervention (if this is the cutoff to which the school/district agreed).

Test Reliability

In assessment, reliability refers to the consistency with which a test measures a particular ability, in this case early mathematics. When discussing reliability, Anastasi and Urbina (1998) discussed several types of reliability analyses that can be conducted. We report analyses on coefficient alpha, alternate forms (immediate), alternate forms (delayed), test-retest, and scorer differences. We also provide data regarding the tests' standard error of measurement.

Alpha

Cronbach's (1951) alpha provides an index of content sampling error. We subjected each of the TEMI-PM and TEMI-O subtests and totals to reliability studies that produced alphas. However, alphas reported for the TEMI-PM should be interpreted with caution; some experts state that timed tests should not be examined using alpha because not all items will be answered

© 2008 The University of Texas System/Texas Education Agency

by all students (Rozeboom, 1966). However, we have found examples in peer-reviewed journal articles of researchers reporting coefficients alpha on timed tests, so it is not an uncommon practice (see Fuchs, Compton, Fuchs, Paulsen, Bryant, & Hamlett, 2005; and Fuchs et al., 2006). To account somewhat for the variability in test items taken by students, we provide two alpha indexes. The first is based on all items of the TEMI-PM subtests and totals. The second reports alphas only for items that at least 75% of the students answered (see Ho Yu, undated). For example, for Magnitude Comparisons in the spring of first grade, there are 64 items. Of the 64 items, 75% or more of the students in the normative sample answered items 1 though 22. Thus, we generated a coefficient alpha for only these items (see Table 4). Note that the 75% figure is arbitrary. When examining the TEMI-PM alphas, understand that they may overestimate the reliability somewhat, although it is impossible to know by how much. Surely, even those coefficients reported at 75% completion should be considered the upper bounds of reliability estimates for content sampling error.

TEMI-O alphas are reported below. When interpreting reliability coefficients, Salvia and Ysseldyke (2001) note that coefficients of at least .70 are acceptable for group tests, but for individual interpretations, coefficients should meet or exceed .80, with coefficients of .90 or higher preferred. Of particular interest are the coefficients associated with the tests' Total Scores, because the Total Scores are recommended for identification and progress-monitoring purposes. Of the 45 coefficients reported for the TEMI-PM, all exceed .80; the same occurs with the TEMI-O. Many exceed .90. Coefficients alpha are reported for the TEMI-PM and TEMI-O scores for the normative sample in Tables 5 through 8.

Table 5. Coefficients Alpha for TEMI-PM Subtests and Totals, Grades K, 1, and 2, Fall

Grade	Test	Total Items	r	Items @ 75%	r
K	MC	64	.97	10	.92
	NID	28	.89	8	.93
	NS	42	.93	3	.82
	QR	70	.96	16	.96
	Avg.		.94		.91
	TOT	204	.98	37	.95
1	MC	64	.96	22	.95
	NS	42	.93	10	.91
	PV	45	.91	7	.78
	ASC	40	.91	7	.79
	Avg.		.93		.86
	TOT	191	.98	46	.94
2	MC	64	.95	29	.95
	NS	42	.87	11	.85
	PV	32	.85	8	.81
	ASC	40	.93	16	.88
	Avg.		.90		.87
	TOT	178	.97	64	.95

© 2008 The University of Texas System/Texas Education Agency

Table 6. Coefficients Alpha for TEMI-PM Subtests and Totals, Grades K, 1, and 2, Winter

Grade	Test	Total Items	r	Items @ 75%	r
K	MC	64	.97	24	.95
	NID	28	.87	11	.87
	NS	42	.94	11	.92
	QR	70	.96	24	.95
	Avg.		.94		.92
	TOT	204	.98	70	.96
1	MC	64	.96	32	.95
	NS	42	.94	16	.93
	PV	45	.91	12	.86
	ASC	40	.93	13	.88
	Avg.		.94		.91
	TOT	191	.98	73	.96
2	MC	64	.95	35	.91
	NS	42	.91	16	.85
	PV	32	.98	10	.78
	ASC	40	.94	23	.88
	Avg.		.95		.86
	TOT	178	.97	84	.93

Table 7. Coefficients Alpha for TEMI-PM Subtests and Totals, Grades K, 1, and 2, Spring

Grade	Test	Total Items	r	Items @ 75%	r
K	MC	64	.97	32	.97
	NID	28	.87	14	.85
	NS	42	.95	16	.93
	QR	70	.96	28	.94
	Avg.		.94		.92
	TOT	204	.98	90	.97
1	MC	64	.96	34	.94
	NS	42	.94	18	.92
	PV	45	.93	16	.92
	ASC	40	.94	16	.90
	Avg.		.94		.92
	TOT	191	.98	84	.96
2	MC	64	.96	41	.93
	NS	42	.93	16	.81
	PV	32	.90	12	.73
	ASC	40	.94	26	.89
	Avg.		.93		.84
	TOT	178	.98	95	.94

© 2008 The University of Texas System/Texas Education Agency

Table 8. Coefficients Alpha for TEMI-O Subtests and Totals, Grades K, 1, and 2, Fall/Winter/Spring

	Fall		Winter		Spring		
Grade/Test	Total Items	r	Total Items	R	Total Items	r	Average
K MPS	35	.87	35	.84	35	.83	.84
1 MPS	38	.80	38	.81	38	.87	.83
C	30	.80	30	.87	30	.87	.85
TOT	68	.86	68	.90	68	.92	.90
2 MPS	40	.77	40	.78	40	.82	.80
C	54	.89	54	.90	54	.91	.90
TOT	94	.90	94	.90	94	.92	.90

Alternate Forms (Immediate)

When more than one form is available, and the forms are administered at the same time, resulting correlation coefficients provide an estimate of content sampling error. For the TEMI-PM, we administered the four forms (A, B, C, and D) in counterbalanced order. On 1 day, two forms of each subtest were administered; on a second day (from 1 day to 1 week later), the other two forms were given. For the forms that were administered on either day, resulting correlation coefficients would provide estimates of content sampling error. Tables 9, 11, and 13 provide means and standard deviations across the three forms that constitute the current TEMI-PM (A, B, and C). Tables 10, 12, and 14 provide the average of the two coefficients (Day 1 and Day 2) for the three forms.

Table 9. Means and Standard Deviations Across Forms for Kindergarten

	MC			NID			NS		
	A	B	C	A	B	C	A	B	C
Mean	35.28	34.82	35.32	12.05	12.20	12.02	14.27	13.68	13.70
SD	16.98	18.01	17.16	4.35	5.24	4.83	8.04	8.47	8.46

	QR			Total		
	A	B	C	A	B	C
Mean	32.06	29.78	29.74	93.11	90.19	90.59
SD	12.43	11.07	10.60	37.34	37.89	36.83

Table 10. Alternate Forms (Immediate) Correlations for Kindergarten

	MC		NID		NS		QR		Total	
	A	B	A	B	A	B	A	B	A	B
A		.85		.83		.85		84		.92
C	.76	.81	.75	.78	.84	.81	.72	.80	.88	.89

© 2008 The University of Texas System/Texas Education Agency

Table 11. Means and Standard Deviations Across Forms for Grade 1

	MC			NS			PV		
	A	B	C	A	B	C	A	B	C
Mean	41.08	38.94	37.77	18.95	19.22	18.51	14.38	15.99	17.81
SD	11.65	11.71	11.64	7.45	7.84	7.31	5.16	5.17	5.49

	ASC			Total		
	A	B	C	A	B	C
Mean	16.64	15.13	16.39	91.43	89.68	90.91
SD	7.82	7.18	7.45	27.14	27.17	27.06

Table 12. Alternate Forms (Immediate) Correlations for Grade 1

	MC		NS		PV		ASC		Total	
	A	B	A	B	A	B	A	B	A	B
A		.76		.80		.76		.81		.89
C	.82	.78	.84	.73	.75	.76	.75	.84	.89	.89

Table 13. Means and Standard Deviations Across Forms for Grade 2

	MC			NS			PV		
	A	B	C	A	B	C	A	B	C
Mean	49.67	50.08	50.48	20.11	21.08	20.38	16.33	14.80	15.41
SD	10.88	11.41	10.72	7.86	7.21	7.32	5.73	5.62	5.91

	ASC			Total		
	A	B	C	A	B	C
Mean	29.24	28.07	28.81	115.62	114.33	115.29
SD	8.56	8.93	8.83	27.77	27.48	27.88

Table 14. Alternate Forms (Immediate) Correlations for Grade 2

	MC		NS		PV		ASC		Total	
	A	B	A	B	A	B	A	B	A	B
A		.81		.88		.80		.84		.93
C	.83	.83	.87	.84	.87	.88	.85	.83	.93	.92

For the TEMI-O, we conducted only one immediate alternate forms study—for Computations Forms A and C. The items were combined into one longer test, and students completed the tests at one time. The Form A items were summed for each student, as were the Form B items; totals were correlated with one another, yielding a reliability coefficient of .97 for grade 1 and .98 for grade 2. For MPS, we administered Forms A and B in counterbalanced order in the winter of 2007. For kindergarten (n = 292), Form A means and standard deviations were 25.6 and 5.7, respectively; for Form B, they were 26.2 and 5.1. The two forms correlated at .76. For grade 1 (n = 621), Form A means and standard deviations, respectively, were 26.2 and 5.1; for Form B, they were 25.6 and 5.5 The two forms correlated at .72. For the second-graders (n = 612), Form A means and standard deviations were 26.4 and 5.5, respectively; for Form B, they were 26.1 and 5.4. The two forms correlated at .71.

© 2008 The University of Texas System/Texas Education Agency

Alternate Forms (Delayed)

As noted above, the TEMI-PM subtests were administered in counterbalanced order. We correlated the results across days, and resulting coefficients are found below. These coefficients contain two sources of error variance: content sampling error and time sampling error.

For the TEMI-O, we administered Form A on 1 day and Form B on another day (from 1 day to 1 week later). Correlating the results provided the coefficients reported in Tables 15 through 17 below. Again, these coefficients contain two sources of error variance: content sampling error and time sampling error.

Table 15. Alternate Forms (Delayed) Correlations for Grade 2

	MC		NID		NS		QR		Total	
	A	B	A	B	A	B	A	B	A	B
A		.62		.68		.74		.63		.75
C	.78	.82	.78	.78	.78	.80	.73	.76	.85	.88

Table 16. Alternate Forms (Delayed) Correlations for Grade 2

	MC		NS		PV		ASC		Total	
	A	B	A	B	A	B	A	B	A	B
A		.72		.77		.67		.75		.83
C	.75	.77	.74	.78	.69	.70	.78	.82	.85	.89

Table 17. Alternate Forms (Delayed) Correlations for Grade 2

	MC		NS		PV		ASC		Total	
	A	B	A	B	A	B	A	B	A	B
A		.80		.80		.81		.80		.88
C	.80	.77	.80	.82	.84	.83	.77	.80	.89	.88

Test-Retest

To assess time sampling error alone, simple calculations are made using the alternate forms (immediate and delayed) results. We have paraphrased Anastasia and Urbina's (1997) formulas:

- A = 1 – (content sampling + time sampling error) (alternate forms, delayed)
- B = 1 – content sampling error (alternate forms, immediate)
- C = A – B
- Time sampling error (test-retest) = 1 – C

So, using Magnitude Comparisons, Grade 2, Forms B and C (see Tables 17 and 14), as an example:

- A = 1 – .77 = .23
- B = 1 – .83 = .17
- C = .23 – .17 = .06
- Time sampling error (test-retest) = 1 – .06 = .94

Results depicting time sampling error alone are found in Tables 18, 19, and 20 below.

© 2008 The University of Texas System/Texas Education Agency

Table 18. Test-Retest Correlations for Kindergarten

	MC		NID		NS		QR		Total	
	A	B	A	B	A	B	A	B	A	B
A		.77		.85		.89		.79		.83
C	NA	NA	NA	NA	.94	.99	NA	.96	.97	.99

Table 19. Test-Retest Correlations for Grade 1

	MC		NS		PV		ASC		Total	
	A	B	A	B	A	B	A	B	A	B
A		.96		.97		.91		.94		.94
C	.93	.99	.90	NA	.84	.94	.NA	.98	.96	.99

Table 20. Test-Retest Correlations for Grade 2

	MC		NS		PV		ASC		Total	
	A	B	A	B	A	B	A	B	A	B
A		.99		.92		.99		.96		.95
C	.97	.94	.93	.98	.97	.95	.92	.97	.96	.96

Scorer Differences

To examine scorer differences, we had two project staff members select 30 protocols from the spring of 2008 at each grade for Form A of the TEMI-PM and TEMI-O. They then hand scored the tests using the Scoring Sheets that are available to the public. Their resulting test scores for the subtests and totals were intercorrelated, and the results reflect the amount of variance that is attributable to scorer differences. As can be seen in Tables 21 through 26, this error variance is almost (and, in some cases, is completely) nonexistent.

Table 21. Inter-scorer Correlations for Kindergarten TEMI-PM

	MC	NID	NS	QR	Total
R	1.00	.999	1.00	.993	1.00

Table 22. Inter-scorer Correlations for Kindergarten TEMI-O

	MPS
R	.997

Table 23. Inter-scorer Correlations for Grade 1 TEMI-PM

	MC	NS	PV	ASC	Total
R	1.00	.998	.994	.997	1.00

Table 24. Inter-scorer Correlations for Grade 1 TEMI-O

	MPS	C	Total
R	.998	.998	.998

Table 25. Inter-scorer Correlations for Grade 2 TEMI-PM

	MC	NS	PV	ASC	Total
R	.996	.999	.995	1.00	1.00

© 2008 The University of Texas System/Texas Education Agency

Table 26. Inter-scorer Correlations for Grade 2 TEMI-O

	MPS	C	Total
R	.998	1.00	1.00

Standard Error of Measurement

The standard error of measurement (SEM) is used to identify a confidence interval surrounding a particular test score. Thus, the SEM reflects an individual's test score due to imperfect test reliability.

The formula for SEM is SEM = SD multiplied by the square root of 1 – *r* (SD = standard deviation; *r* = reliability). The clinical value of SEMs is demonstrated by a first-grade student who scored 84 points on the TEMI-PM Total Score in the spring. By applying the SEM, the teacher knows with 68% probability that the "true score" is between 77.8 and 90.2, with 95% probability that the "true score" is between 71.9 and 96.2 (1.96 * 6.2), and with 99% probability that the "true score" is between 68 and 100 (2.58 * 6.2).

When making intervention decisions, it is helpful to consider the SEM and the student's "true score," especially when there is doubt about whether the student should qualify for intervention. For example, a teacher may rate the student as having average math skills. The TEMI-PM Total Score may place the child at the 24th percentile, therefore just barely qualifying for intervention. Consideration of the "true score" might lead to a decision that the student should not qualify and that instead, the student should continue with the Tier I program only, with the teacher keeping a close eye on the student's progress. Standard errors of measurement are found below in Tables 27 through 32.

Table 27. Standard Errors of Measurement for Kindergarten TEMI-PM

	Fall			Winter			Spring		
	.68	.95	.99	.68	.95	.99	.68	.95	.99
TEMI-PM MC	3.8	7.4	9.8	3.3	6.5	8.5	2.6	5.1	6.7
TEMI-PM NID	1.2	2.4	3.1	1.4	2.7	3.6	1.6	3.1	4.1
TEMI-PM NS	2.9	5.7	7.5	2.2	4.3	5.7	2.2	4.3	5.7
TEMI-PM QR	2.3	4.5	5.9	2.4	4.7	6.2	2.6	5.1	6.7
TEMI-PM Total	7.0	13.7	18.1	6.2	12.2	16.0	5.6	11.0	14.4

Table 28. Standard Errors of Measurement for Grade 1 TEMI-PM

	Fall			Winter			Spring		
	.68	.95	.99	.68	.95	.99	.68	.95	.99
TEMI-PM MC	2.6	5.1	6.7	2.3	4.5	5.9	2.7	5.3	7.0
TEMI-PM NS	1.9	3.7	4.9	1.9	3.7	4.9	2.1	4.1	5.4
TEMI-PM PV	1.8	3.5	4.6	1.9	3.7	4.9	1.6	3.1	4.1
TEMI-PM ASC	2.1	4.1	5.4	2.4	4.7	6.2	2.4	4.7	6.2
TEMI-PM Total	5.4	10.6	13.9	5.1	10.0	13.2	5.4	10.6	13.9

© 2008 The University of Texas System/Texas Education Agency

Table 29. Standard Errors of Measurement for Grade 2 TEMI-PM

	Fall			Winter			Spring		
	.68	.95	.99	.68	.95	.99	.68	.95	.99
TEMI-PM MC	2.3	4.5	5.9	3.1	6.1	8.0	2.9	5.7	7.5
TEMI-PM NS	1.6	3.1	4.1	2.2	4.3	5.7	2.9	5.7	7.5
TEMI-PM PV	1.7	3.3	4.4	2.3	4.5	5.9	2.8	5.5	7.2
TEMI-PM ASC	2.6	5.1	6.7	2.7	5.3	7.0	2.7	5.3	7.0
TEMI-PM Total	4.6	9.0	11.9	6.1	12.0	15.7	6.2	12.2	16.0

Table 30. Standard Errors of Measurement for Kindergarten TEMI-O

	Fall			Winter			Spring		
	.68	.95	.99	.68	.95	.99	.68	.95	.99
TEMI-O MPS	2.4	4.7	6.2	2.2	4.3	5.7	2.1	4.1	5.4

Table 31. Standard Errors of Measurement for Grade 1 TEMI-O

	Fall			Winter			Spring		
	.68	.95	.99	.68	.95	.99	.68	.95	.99
TEMI-O MPS	2.3	4.5	5.9	2.3	4.5	5.9	2.1	4.1	5.4
TEMI-O C	2.1	4.1	5.4	2.1	4.1	5.4	2.1	4.1	5.4
TEMI-O Total	3.3	6.5	8.5	3.2	6.3	8.3	3.1	6.0	7.9

Table 32. Standard Errors of Measurement for Grade 2 TEMI-O

	Fall			Winter			Spring		
	.68	.95	.99	.68	.95	.99	.68	.95	.99
TEMI-O MPS	2.8	5.5	7.2	2.4	4.7	6.2	2.2	4.3	5.7
TEMI-O C	2.7	5.3	7.0	2.4	4.7	6.2	2.3	4.5	5.9
TEMI-O Total	3.7	7.3	9.5	3.8	7.4	9.8	3.6	7.1	9.3

Reliability Summary

In this section of the manual, we have provided evidence for several types of reliability. When the data are considered in their entirety, there is ample evidence that the TEMI-PM and TEMI-O are reliable measures, especially when the Total Score is used.

Validity

In its most basic terms, tests are said to be valid if they demonstrate that they measure what they are supposed to measure. But it is easier to define validity than it is to demonstrate that a particular test has valid scores. This is because validity is relative instead of being an absolute concept. The validity of a test's scores will vary according to the purpose for which the scores are used. Therefore, a test's validity must be examined and re-examined until a thorough body of research is accumulated. Simply put, the study of any test's validity is a lengthy and ongoing process. That said, test authors must provide a "starter set" of data that supports the validity of a test's scores for specific purposes. That is what we do in this section of the manual.

A number of testing experts—for example, Aiken (2000), Anastasi and Urbina (1997), Linn and Gronlund (1995), and Salvia and Ysseldyke (2001)—suggest that test developers

© 2008 The University of Texas System/Texas Education Agency

should examine different types of validity. We have chosen to follow Anastasi and Urbina's triad of validity (1997): content-description, criterion-prediction, and construct-identification.

Content-Description Validity

Anastasi and Urbina (1997) describe content-description validation procedures as involving "the systematic examination of the test content to determine whether it covers a representative sample of the behavior domain to be measured" (pp. 114–115). There is little doubt that this kind of validity must be built into the test at the time that subtests are conceptualized and items developed. Test authors usually deal with content-description validity by showing that the abilities chosen to be measured are consistent with current knowledge about the content that is being measured and that the items hold up empirically. Three demonstrations of content-descriptive validity are offered for the TEMI-PM and the TEMI-O. First, the subtests of each measure are described and their rationale presented. Second, traditional item analyses data are presented. Finally, the validity of the test is supported by the results of differential item functioning analysis, which shows the absence of test bias.

TEMI-PM and TEMI-O Format and Content Rationale

Before we describe the subtests and provide their rationale, we should address the formats of the tests, specifically why we deiced to use group testing formats and why we incorporate timed tests for the TEMI-PM. We initially created and field tested individual tests for use in this project. However, the amount of time it took to administer and score the tests was prohibitive. Because the tests would eventually be administered by classroom teachers, we were concerned about the amount of instructional time that would be lost. With a group format, the test takes about 45 minutes on each of 2 days and about 30 minutes the third day. That totals 2 hours of lost instructional time three times a year. For a teacher to test 20 students individually, the amount of instructional time lost could be as much as 40 hours (2 hours multiplied by 20 students), which is clearly unacceptable. Even if the test were administered more quickly, there is still the chance that 20 hours or so would be spent on testing, a much greater investment of time than the 2 hours in the current format.

A consultant to the project reinforced our idea, stating that a recurring issue in Reading First assessments was the amount of testing time required. If he had his druthers, all tests would be group administered, provided they could be reliable tests that yield valid results. That was our challenge in building the group measures.

The second issue involves the inclusion of timed tests for the TEMI-PM. We wanted to replicate, to some extent, tests that were being currently used for progress-monitoring purposes. Many of the curriculum-based measurement tests created by the Curriculum-Based Measurement Warehouse (www.interventioncentral.org/htmdocs/interventions/cbmwarehouse.php) are timed measures, as are many described by the Research Institute on Progress Monitoring (www.progressmonitoring.net/Techreports/TREarlymath.doc) and the National Center on Student Progress Monitoring (www.studentprogress.org). We also found a study conducted by Fuchs, Fuchs, Mathes, and Lipsey (2002) that addressed timed reading tests. These researchers noted that:

> …regarding [effect sizes] ESs for timed measurements, whereby students were required to perform (i.e., read aloud, read silently, answer questions, match words to meanings) within a fixed time, the ESs increased to well beyond one full standard deviation unit. And, in a similar way, when LD and LA samples had been identified using data-based methods, the overall ES of 0.61 rose to beyond a full standard deviation unit. Findings,

© 2008 The University of Texas System/Texas Education Agency

> therefore, suggest that researchers and school personnel in fact do identify as LD those children who have appreciably more severe reading problems compared to other low-performing students who go unidentified (pp. 94–95).

We decided to generalize their findings to our mathematics progress-monitoring measures, supported by Berch's (2005) assertion that "recent evidence from studies of school-age children with MD suggests that it may be crucial to measure the speed of executing such quantity discriminations, as this variable can reveal subtle yet important differences in numerical information processing that may not be tapped by assessing accuracy alone" (p. 355).

For the TEMI-O, we selected items for Mathematics Problem Solving by examining the Texas Essential Knowledge and Skills (TEKS) and leading basal mathematics series to see how the TEKS were being taught. The TEKS provides teachers and administrators with essential skills that should be taught in six key areas: number, operations, and quantitative reasoning; pattern, relationships, and algebraic thinking; geometry and spatial reasoning; measurement; probability and statistics; and underlying processes and mathematical tools. These areas are similar to those offered by the National Council of Teachers of Mathematics (NCTM). We examined basal mathematics textbooks (kindergarten, grade 1, and grade 2) that were on the state adoption list (in 2005, when we developed the test) and identified five that were among the most widely used in the state (using data supplied by the state): Harcourt, Scott Foresman-Addison Wesley, SRA, Everyday Mathematics, and Houghton Mifflin. Our goal was to identify how the texts taught the TEKS and assessed the skills that were taught. After the items were created, we had a mathematics educator who was a member of our project staff examine the items with respect to the TEKS. Finally, we asked for input from the dozens of teachers who participated in the standardization of the instrument.

Even with the care we took to create the TEMI-O MPS items, before using the TEMI-O, school officials should compare the content of the MPS to the scope and sequence of the skills being taught in the district. If there is not a match to what is being taught, the TEMI-O MPS may not yield results that are valid indicators of an individual students' response to general education instruction.

For the TEMI-O Computation, we examined scope-and-sequence charts of the basals and also scope-and-sequence charts for addition, subtraction, multiplication, and division in grades 1 and 2. Users should be cautioned about making judgments of student abilities in each area, because only a single item usually represents the item type (e.g., addition doubles, make tens plus more). However, collectively, the subtest provides a comprehensive assessment of whole-number computation at the early grades. Again, districts should check their programs' scope and sequence to determine the skill correspondence.

Having presented our rationale for group administration and the use of timed tests, we now provide the rationale for each subtest of the TEMI-PM and TEMI-O. For each subtest, we state what is measured, how the content is assessed, what students are required to do, and precedents that exist.

TEMI-PM Magnitude Comparisons (Kindergarten, Grade 1, and Grade 2)

- **What is measured:** Magnitude Comparisons involves a variety of skills: numeral recognition and knowledge (0 through 20 in kindergarten, 0 through 99 in grade 1, 0 through 999 in grade 2); quantity recognition; place value; "same as, less than, greater than"; and the ability to handle a pencil in order to draw a circle or loop, turn pages of a booklet, and listen to and follow directions. Also, students must sustain attention and be able to track across columns and avoid skipping rows as they work.

© 2008 The University of Texas System/Texas Education Agency

- **How the content is measured:** Students look at two numbers that appear side-by-side in a box in their student booklet (a vertical dotted line separates the two numbers). As a fluency measure, the test is designed to see how many items the student can answer correctly in 2 minutes by circling the bigger of the two numbers (in kindergarten; first-graders circle the smaller number; second-graders circle the number "that is less") or circling both numbers if they are the same (equal).
- **What students do:** Students must pay attention, listen carefully, and do their best to look at two numbers side-by-side and circle the number that is their answer or circle both numbers if they are the same. Students should start on time (not early) and stop testing and put their pencil down when the examiner says, "Stop."
- **Precedent:** This task has a history of use in neuropsychological and educational assessments. Dehaene (1997) described several studies where individuals are asked to compare the size of numbers. Also, quantity discrimination tests are available online (www.interventioncentral.org/htmdocs/interventions/cbmwarehouse.php), and Clarke and Shinn (2004) have written their *Quantity Discrimination Measure Verbal 1–20*.

TEMI-PM Number Identification (Kindergarten Only)

- **What is measured:** Number Identification involves a variety of skills: numeral recognition and knowledge (1–20), place value, and sequence.
- **How the content is measured:** Students look at rows of squares, up to five squares per row, and count the number of squares they see. Students then look at four possible response choices, next the stimulus boxes, and circle their answer. As a fluency measure, the test is designed to see how many items students can answer correctly in 2 minutes.
- **What students do:** Students should pay attention, listen carefully, do their best to count the squares, look at the response choices, and circle the response choice that shows how many squares there are in all. We also ask them to start on time (not early) and stop testing and put their pencil down when the examiner says, "Stop." Children must manage a pencil in order to draw a circle or loop, turn pages of a booklet, and listen to and follow directions. Also, students must pay sustained attention while being able to track across columns and not skip rows as they work.
- **Precedent:** Number identification tasks have appeared on many early mathematics tests (e.g., KeyMath Diagnostic Arithmetic Test, Woodcock-Johnson Revised, and Clarke and Shinn's [2004] *Number Identification Measure 1–20*).

TEMI-PM Number Sequences (Kindergarten, Grade 1, and Grade 2)

- **What is measured:** Number Sequences imbeds several skills: numeral recognition and knowledge, counting ahead or counting back (0 through 20 in kindergarten, 0 through 99 in grade 1, 0 through 999 in grade 2), place value, and "less than and greater than."
- **How the content is measured:** Students look at a three-number sequence; one number of the sequence is missing and is represented by a blank. The missing number may be the first number in the sequence, the second number, or the third number. In kindergarten and grade 1, the student then looks at four possible response choices, in boxes below the stimulus series, and circles the answer. In grade 2, students write the number on the blank (representing the missing number). As a fluency measure, the test is designed to see how many items the student can answer correctly in 2 minutes.
- **What students do:** Students are asked to pay attention, listen carefully, and do their best to look at the series and the response choices and circle the response choice (or, for

© 2008 The University of Texas System/Texas Education Agency

second-graders, write the number) that completes the three-number series. We also ask them to start on time, not early, and stop testing and put their pencil down when the examiner says, "Stop." Students also are asked to manage a pencil in order to draw a circle or loop (kindergarten and grade 1) or write numerals to 999 (grade 2), turn pages of a booklet, and to listen to and follow directions. Finally, students must pay sustained attention and be able to track across columns and not skip rows as they work.

- **Precedent:** The skill of identifying missing numbers in a sequence is often found on tests that assess number sense. Also, missing number tests are available online (www.interventioncentral.org/htmdocs/interventions/cbmwarehouse.php), and Clarke and Shinn (2004) include *Missing Number Measure Blank Varied 1–20* as one of their early mathematics screening measures.

TEMI-PM Quantity Recognition (Kindergarten Only)

- **What is measured:** Quantity Recognition involves a variety of skills, including the ability to: see and recognize numerals, recognize quantity immediately (a skill known as subitizing), or count ahead to 6.
- **How the content is measured:** Students look at randomly placed dots clustered near one another, ranging from 1 to 6 dots. The student then looks at six possible response choices: 1, 2, 3, 4, 5, or 6, in boxes below the stimulus item, and circles the number that corresponds to the number of dots shown. As a fluency measure, the test is designed to see how many items the student can answer correctly in 2 minutes.
- **What students do:** We ask students to pay attention, listen carefully, do their best to recognize quantity immediately or count the circles, look at the response choices, and circle the response choice that shows how many dots they see. We also ask them to start on time, not early, and stop testing and put their pencil down when the examiner says, "Stop." Children must manage a pencil in order to draw a circle or loop, turn pages of a booklet, and listen to and follow directions. Also, students must stay on task, be able to track across columns, and not skip rows as they work.
- **Precedent:** Kaufman, Lord, Reese, and Volkmann (1949) used the term *subitizing* to describe the rapid, correct, and self-assured judgment of the quantity represented by small numbers of items. Several researchers have used the format for this test, including Benoit, Lehalle, and Jouen (2004). These researchers differentiated "familiar" from "unfamiliar" dot configurations. Familiar configurations are similar to those used on dice; unfamiliar configurations are more random. We selected unfamiliar configurations, because we wanted to reduce the effects of board game influence (i.e., we posited that students who played board games as preschoolers may have an advantage over those who did not, and we wanted subitizing to be assessed, not board game experience).

TEMI-PM Place Value (Grades 1 and 2 Only)

- **What is measured:** Place Value involves several skills, including the ability to: see the stacks of tens and ones (grade 1) or hundreds, tens, and ones (grade 2); recognize the quantity represented by the stacks; recognize numerals (four response choices for each item in grade 1); or write a numeral (grade 2).
- **How the content is measured:** Students look at stacks of tens and ones up to 99 (grade 1) or hundreds, tens, and ones up to 999 (grade 2). First-graders then look at four possible response choices, in boxes below the stimulus item, and circle the number that shows "how many." Second-graders identify "how many" and write their answer in a box next

© 2008 The University of Texas System/Texas Education Agency

to the stacks. As a fluency measure, the test is designed to see how many items students can answer correctly in 2 minutes.

- **What students do:** We ask students to pay attention, listen carefully, count by tens and by ones (or count by hundreds, tens, and ones), look at the response choices, and circle the response choice (grade 1) or write the number (grade 2) that shows "how many" they see. We also ask them to start on time, not early, and stop testing and put their pencil down when the examiner says, "Stop." Students must manipulate a pencil in order to draw a circle or loop around the response choice (grade 1) or to write a numeral (grade 2), turn pages of a booklet, and listen to and follow directions. Attention and tracking skills also come into play.
- **Precedent:** Place value and base 10 models are taught in a variety of ways in early mathematics basals. In end-of-chapter and review tests, students are often asked to look at stacks of ones, tens, and sometimes hundreds (in later grades) and identify the number that is represented.

TEMI-PM Addition/Subtraction Combinations (Grades 1 and 2 Only)

- **What is measured:** Addition/Subtraction Combinations assess basic math facts and involves several skills: recognizing numerals, recognizing operational signs (plus and minus), and computing (addition and subtraction).
- **How the content is measured:** Students look at addition and subtraction problems on a page and then compute and write the answer to each problem. As a fluency measure, the test is designed to see how many items the student can answer correctly in 2 minutes.
- **What students do:** We ask students to pay attention, listen carefully, and do their best to compute and write the answer. We also ask them to start on time, not early, and stop testing and put their pencil down when the examiner says, "Stop." Children are asked to use a pencil in order to write a numeral (the answer—both grades 1 and 2). They must have the ability to turn pages of a booklet and listen to and follow directions.
- **Precedent:** Most tests of mathematics have measures similar to this test. We developed this test based on the content of the Computational Arithmetic Program (Smith & Lovitt, 1982). Tests that use this format include the AIMSweb *Math Facts* (www.aimsweb.com/measures/math/scoring.php) and a variety of curriculum-based assessment math computation probes (www.interventioncentral.org/htmdocs/tools/mathprobe/addsing.php).

TEMI-O Mathematics Problem Solving (K, 1, and 2)

- **What is measured:** Mathematics Problem Solving assesses children's ability to listen and pay attention as items are read aloud, access their mathematics knowledge, examine the stimulus prompt, survey the response choices, and connect two dots to mark the answer. Students who do well on Mathematics Problem Solving exhibit skills across the TEKS. Students who do poorly may have difficulty with all of the TEKS or may exhibit strengths and struggles across abilities.
- **How the content is measured:** Students solve three practice items, one at a time. The first item is designed to show students how to mark their answer sheet by connecting two dots. Remaining practice items are designed to familiarize students with the format of the test. Teachers read aloud each test item as students follow along in their test booklets. Teachers proceed briskly through the test, allowing only about 10 seconds for the students to respond after the item has been read to them. Students are given more time for

© 2008 The University of Texas System/Texas Education Agency

some items on the test; these items are marked on the instructions. Pacing is important. Teachers do not wait for all students to mark their booklets; instead, teachers proceed in a brisk fashion. Typically, allowing long periods of time to pass between items frustrates the stronger students and does little to help struggling students.

- **What students do:** Students have to listen carefully as the teacher reads the items because the items are not repeated in their entirety. Students also have to keep their place in the booklets to ensure they mark the appropriate item. Students view stimulus items and response choices and must be able to connect two dots to mark their answer. Students also must be able to turn pages with some degree of facility and make sure they are on the right page.
- **Precedent:** Numerous tests examine broad-based math skills. The SAT-10, the ITBS, and several other group-administered measures have mathematics problem solving tests.

TEMI-O Computation

- **What is measured:** Computation assesses arithmetic calculation and involves several skills: recognizing numerals and operational signs (plus and minus for grade 1); computing (addition and subtraction for grade 1); and the ability to manage a pencil in order to circle the answer, turn pages of a booklet, and listen to and follow directions.
- **How the content is measured:** Students look at arithmetic problems on a page and then compute and circle the answer to each problem.
- **What students do:** We ask students to pay attention, listen carefully, and do their best to compute and circle the answer to each problem.
- **Precedent:** Numerous tests examine math computation skills. The SAT-10, ITBS, and several other group-administered measures have mathematics problem solving tests. In addition, tests like the KeyMath, WJ-R, WRAT-3, and Diagnostic Achievement Battery – 4th edition (Newcomer, 2008) all have computation subtests.

Conventional Item Analysis

In the previous sections, we described the subtests of the TEMI-O and TEMI-PM and provided the rationale for their inclusion in our test battery. Here we provide the results of empirical studies of content-description validity. Specifically, we conducted traditional item analysis to identify each item's discriminating power. Discriminating power, or item validity, as it is sometimes called, is defined as "the degree to which an item differentiates correctly among test takers in the behavior that the test is designed to measure" (Anastasi & Urbina, 1997, p. 179). The discriminating power index is a correlation coefficient depicting the relationship between a particular item and the other items on the test.

Hammill and Bryant (2005) identified that more than 50 different indexes of item discrimination have been developed for use in test construction, although they all provide similar results (Anastasi & Urbina, 1997; Guilford & Fruchter, 1978; Oosterhof, 1976). Ebel (1972) and Pyrczak (1973) suggested that discrimination indexes of .35 or higher are acceptable. Anastasi and Urbina (1997) and Garrett (1965) pointed out that indexes as low as .20 are acceptable under some circumstances.

To examine the item characteristics of the TEMI-PM and TEMI-O subtests, an item analysis was undertaken using three samples. The first sample is from research we conducted in Pflugerville Independent School District (PISD). The second sample is from intervention scale-up sites in Lubbock-Cooper ISD and Copperas Cove ISD. And the third sample is from the normative sample from more than 60 school districts in Texas. The resulting discriminating

© 2008 The University of Texas System/Texas Education Agency

power coefficients (corrected for part-whole effect) are reported below in Tables 34 through 37. We provide two sets of data for the TEMI-PM; the first set is composed of items where at least 75% of the students answered the items (correctly or incorrectly—see the discussion of coefficients alpha in the Reliability section of the manual), and the second set of data is on all items. For the TEMI-O, data are provided for all items, because the TEMI-O subtests are not timed.

Table 34. Median and Range of Kindergarten, Grade 1, and Grade 2 TEMI-O Discriminating Powers—Fall, Winter, and Spring

	Fall		Winter		Spring	
Grade/Test	Mdn Disc P	Range	Mdn Disc P	Range	Mdn Disc P	Range
K MPS	.36	.05–.50	.35	.10–.46	.37	-.02–.50
1 MPS	.29	-.05–.45	.34	-.08–.51	.42	-.03–.56
1 C	.36	-.03–.48	.40	.13–.56	.44	.22–.54
1 TOT	.29	-.10–.47	.35	-.09–.53	.40	-.02–.59
2 MPS	.28	.03–.44	.28	.00–.45	.31	.01–.42
2 C	.35	-.02–.47	.42	.00–.53	.39	.19–.54
2 TOT	.27	-.04–.41	.31	-.02–.49	.32	-.04–.51

Table 35. Median and Range of Fall Kindergarten, Grade 1, and Grade 2 TEMI-PM Discriminating Powers

Grade	Test	Total Mdn	Range	75% Mdn	Range
K	MC	.58	.24–.78	.72	.49–.80
	NID	.37	-.03–.73	.77	.67–.82
	NS	.41	.12–.73	.68	.63–.74
	QR	.53	.23–.72	.79	.67–.84
1	MC	.51	.27–.70	.69	.45–.75
	NS	.46	.32–.64	.67	.58–.78
	PV	.54	.16–.63	.49	.42–.65
	ASC	.49	.33–.61	.52	.47–.54
2	MC	.48	.17–.65	.69	.44–.79
	NS	.35	-.06–.61	.61	.31–.71
	PV	.37	-.05–.61	.61	.34–.69
	ASC	.48	.24–.73	.53	.45–.68

© 2008 The University of Texas System/Texas Education Agency

Table 36. Median and Range of Winter Kindergarten, Grade 1, and Grade 2 TEMI-PM Discriminating Powers

Grade	Test	Total Mdn	Range	75% Mdn	Range
K	MC	.57	.32–.74	.67	.44–.78
	NID	.42	.32–.56	.67	.44–.76
	NS	.49	.14–.70	.68	.60–.78
	QR	.52	.28–.67	.68	.53–.75
1	MC	.50	.29–.69	.61	.42–.77
	NS	.50	.28–.70	.67	.52–.73
	PV	.41	.12–.67	.54	.26–.74
	ASC	.44	.19–.69	.55	.41–.73
2	MC	.39	.02–.75	.55	.10–.63
	NS	.43	.18–.67	.50	.30–.64
	PV	.40	.11–.72	.49	.26–.55
	ASC	.52	.18–.73	.46	.27–.69

Table 37. Median and Range of Spring Kindergarten, Grade 1, and Grade 2 TEMI-PM Discriminating Powers

Grade	Test	Total Mdn	Range	75% Mdn	Range
K	MC	.61	.37–.76	.68	.54–.80
	NID	.41	.28–.56	.51	.41–.60
	NS	.53	.34–.72	.63	.53–.74
	QR	.54	.22–.68	.59	.44–.74
1	MC	.48	.17–.71	.57	.30–.70
	NS	.53	.25–.68	.62	.42–.72
	PV	.45	.27–.66	.61	.48–.74
	ASC	.51	.27–.72	.58	.49–.70
2	MC	.47	.11–.75	.45	.17–.78
	NS	.47	.12–.70	.41	.30–.63
	PV	.50	.08–.71	.40	.19–.51
	ASC	.58	.15–.80	.45	.27–.72

Differential Item Functioning Analysis

The purpose of inspecting for differential item functioning (DIF) is to identify whether certain items are equally as challenging for different subgroups within the population. The reason these analyses are conducted is to protect students against items that may be unfairly biased against them (Camilli & Shepard, 1994). Researchers search for DIF by looking at whether all students with the same ability have equal odds of getting an item correct regardless of their sex, race, or ethnicity. If they do not, then DIF may be responsible.

We used Swaminathan and Rogers' (1990) logistic regression procedure to examine DIF. Multiple comparisons were made for each of the group pairings (male vs. female, African American vs. non-African American, and Hispanic American vs. non-Hispanic American). We

© 2008 The University of Texas System/Texas Education Agency

followed Miller and Spray's (1993) recommendation that a significance level of .001 be chosen, in order to prevent the overidentification of potentially biased items (i.e., a high Type I error rate) that can occur when large numbers of comparisons are made.

We also examined effect size to evaluate the magnitude or amount of DIF present in the tests. We followed Zumbo's (1999) suggestion and used Nagelkerke's R^2 (a weighted least squares effect size) to determine the degree of each item's DIF. Cohen's (1992) conventions for small, medium, and large effects led Jodoin and Gierl (2000) to suggest that R^2 less than 0.035 is indicative of negligible DIF, R^2 greater than 0.034 but less than 0.070 demonstrates moderate DIF, and R^2 greater than 0.069 is indicative of large DIF.

DIF analyses were conducted on all of the items contained within Forms A, B, and C of the TEMI-O. We found no cases of DIF in any of the items contained within any of the TEMI-O forms at any grade level.

Because the TEMI-PM is a timed test, we elected to look for DIF in items that at least 75% of the students attempted. The results of our analyses indicated that DIF did exist, at least to a minimal extent, in 26 TEMI-PM items (that is, of the 840 items across the three grades, only 26, or .03% [less than one-tenth of 1%], were indicative of bias). Of those 26 items, 10 were found to have a moderate effect of DIF and 14 were found to have a large effect of DIF. The remaining two had only a negligible amount of DIF effect.

Criterion-Prediction Validity

Anastasi and Urbina (1997) note that criterion-prediction procedures "indicate the effectiveness of a test in predicting an individual's performance in specific activities" (p. 118). This form of validity may be either concurrent or predictive, depending on the amount of time lapsed between the administration of the criterion test and the test being validated. When one test is given immediately after the other, it provides evidence of concurrent criterion-prediction validity. When there is a space of time between administrations of the tests, the study provides evidence of predictive criterion-prediction validity.

Concurrent Criterion-Prediction Validity

To examine the concurrent criterion-prediction validity of the TEMI-O and TEMI-PM, we conducted several studies. First, during the fall and spring of the 2006–2007 academic year, we administered Forms A, B, and C of the TEMI-PM and the mathematics subtests of the Stanford Achievement Test – 10th edition (SAT-10; Harcourt Assessment, 2003) to students in kindergarten, grade 1, and grade 2. In the fall of 2006, students in kindergarten were administered the Stanford Early School Achievement Test (SESAT)-1 level, students in grade 1 were administered the SESAT-2 level, and students in grade 2 were given the Primary 1 level. In the spring, kindergarten students were administered the SESAT-2 level, first-grade students were given the Primary-1 level, and second-grade students took the Primary 2 level. The SESAT 1 and 2 have only one math test, Mathematics Problem Solving. The Primary 1 and 2 levels have the Mathematics Problem Solving test and also a Mathematics Procedures test, and also provide a Total Score, which is a composite of the two contributing subtests. For the SAT-10, we were able to compute standard score standard deviations. In all cases, the standard deviations were below 15, signifying a restricted range, so we applied the correction formula described by Guilford and Fruchter (1978) to the achieved coefficients. Two levels of relationship are reported in the tables, uncorrected and corrected.

© 2008 The University of Texas System/Texas Education Agency

In the fall of 2006, students who were administered the TEMI-PM and the TEMI-O also were administered the Iowa Test of Basic Skills (10th edition; ITBS). Students in kindergarten were administered Level 5, students in grade 1 were administered Level 6, and students in grade 2 were given Level 7. By comparing the ITBS mathematics scores to the TEMI-PM and TEMI-O scores, we have additional evidence supporting the concurrent criterion-prediction validity of our math measures.

Also in the fall of 2006, we asked classroom teachers to rate their students' math abilities on a 5-point scale (1 being Poor and 5 being Superior). We correlated the ratings with the TEMI-PM and TEMI-O scores to provide yet another index of concurrent criterion-prediction validity. We asked kindergarten and first-grade teachers to complete a longer Rating Scale in the spring of 2007. Here, we asked them to rate their students' performance in each of the areas of the TEKS. We also asked second-grade teachers to complete the Learning Disabilities Diagnostic Inventory (LDDI; Hammill & Bryant, 1998). Correlating their ratings on both measures to spring performance on the TEMI-PM and TEMI-O provides yet another estimate of the concurrent criterion-prediction validity of the math tests.

In the winter of 2007, we gave the three forms of the TEMI-PM and two forms (A and C) of the TEMI-O. Comparing these students' SAT-10 fall scores to their winter TEMI-PM and TEMI-O scores provides another index of concurrent criterion-prediction validity.

In the spring of 2008, we also gave the SAT-10 to the students. As in the spring of 2007, kindergarten students were administered the SESAT-2 level, first-grade students were given the Primary-1 level, and second-grade students took the Primary 2 level. The results of these studies are depicted below. Note that when we report scores for the SAT-10 and ITBS, we provide two coefficients, one of which reporting the relationship between the measures after accounting for the restricted range of the students who participated in the study. That is, the standard deviations for the SAT-10 scores and ITBS scores were less than 15, the standard deviation for the SAT-10 and ITBS population; thus, the uncorrected correlation is spuriously low. We corrected for the restricted range using a formula provided by Guilford and Fruchter (1978), and these corrected coefficients are reported next to the uncorrected coefficients.

Correlations among the tests yielded the results found in Tables 38 through 68 (the TEMI-PM and TEMI-O with the SAT-10). Tables 69 through 80 provide the results of the TEMI-PM and TEMI-O correlations with the ITBS. And Tables 81 through 88 include the correlation coefficients signifying the relationship between the TEMI-O and TEMI-PM and teacher ratings. Teachers rated their kindergarten and first-grade students as being on a range of 1 (Very Poor) to 5 (Very Superior). For grade 2 students, teachers provided ratings on the LDDI (Hammill & Bryant, 1998). Table titles include the dates of test administration.

Table 38. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Kindergarten TEMI-PM Scores (Fall 2006 Administration) and SAT-10 SESAT-1 Scores (Fall 2006 Administration)

	R with SESAT-1					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.57	.54	.57	.64	.61	.64
TEMI-PM NI	.51	.51	.59	.52	.58	.67
TEMI-PM NS	.62	.63	.60	.69	.69	.67
TEMI-PM QR	.46	.44	.51	.52	.50	.58
TEMI-PM Total	.65	.62	.65	.71	.69	.71

© 2008 The University of Texas System/Texas Education Agency

Table 39. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Kindergarten TEMI-PM Scores (Winter 2007 Administration) and SAT-10 SESAT-1 Scores (Fall 2006 Administration)

	R with SESAT-1					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.47	.44	.44	.54	.50	.50
TEMI-PM NI	.43	.41	.44	.49	.47	.50
TEMI-PM NS	.46	.52	.49	.52	.59	
TEMI-PM QR	.33	.44	.37	.38	.50	.43
TEMI-PM Total	.50	.53	.50	.57	.60	.57

Table 40. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Kindergarten TEMI-PM Scores (Spring 2007 Administration) and SAT-10 SESAT-2 Scores (Spring 2007 Administration)

	R with SESAT-2					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.39	.38	.41	.44	.43	.46
TEMI-PM NI	.45	.46	.53	.51	.52	.59
TEMI-PM NS	.43	.44	.45	.48	.50	.51
TEMI-PM QR	.30	.32	.33	.34	.37	.38
TEMI-PM Total	.45	.45	.49	.51	.51	.55

Table 41. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Kindergarten TEMI-O Scores and SAT-10 SESAT-2 Scores—Spring 2008 Administration

	R with SESAT-2	
TEMI-PM Scores	Uncorrected Form C	Corrected Form C
TEMI-PM MC	.53	.63
TEMI-PM NI	.49	.59
TEMI-PM NS	.47	.57
TEMI-PM QR	.35	.44
TEMI-PM Total	.65	.74

Table 42. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Kindergarten TEMI-O Scores (Fall 2007 Administration) and SAT-10 SESAT-1 Scores (Fall 2007 Administration)

	Uncorrected	Corrected
SAT-10 Score	TEMI-O MPS Form A	TEMI-O MPS Form A
SESAT-1	.75	.79

© 2008 The University of Texas System/Texas Education Agency

Table 43. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Kindergarten TEMI-O Scores (Winter 2007 Administration) and SAT-10 SESAT-1 Scores (Fall 2007 Administration)

	Uncorrected		Corrected	
SAT-10 Score	TEMI-O MPS Form A	TEMI-O MPS Form B	TEMI-O C Form A	TEMI-O MPS Form B
SESAT-1	.69	.69	.73	.73

Table 44. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-O Scores and SAT-10 Primary-1 Scores – Spring 2007 Administration

	Uncorrected		Corrected	
SAT-10 Score	TEMI-O MPS Form A	TEMI-O C Form B	TEMI-O MPS Form A	TEMI-O C Form B
Primary-1 MPS	.70	.35	.73	.38
Primary-1 MP	.59	.34	.63	.37
Primary-1 Total	.69	.37	.72	.40

Table 45. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-PM Scores (Fall 2006 Administration) and SAT-10 SESAT-2 (Fall 2006 Administration)

	R with SESAT-2					
	Uncorrected			Corrected		
TEMI-PM Scores	TEMI-PM Form A	TEMI-PM Form B	TEMI-PM Form C	TEMI-PM Form A	TEMI-PM Form B	TEMI-PM Form C
TEMI-PM MC	.51	.48	.48	.55	.52	.52
TEMI-PM NS	.58	.55	.57	.65	.62	.64
TEMI-PM PV	.58	.59	.56	.62	.63	.60
TEMI-PM ASC	.47	.49	.59	.54	.56	.66
TEMI-PM Total	.66	.64	.64	.72	.70	.70

Table 46. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-PM Scores (Winter 2007 administration) and SAT-10 SESAT-2 Scores (Fall 2006 Administration)

	R with SESAT-2					
	Uncorrected			Corrected		
TEMI-PM Scores	TEMI-PM Form A	TEMI-PM Form B	TEMI-PM Form C	TEMI-PM Form A	TEMI-PM Form B	TEMI-PM Form C
TEMI-PM MC	.51	.48	.48	.55	.52	.52
TEMI-PM NS	.43	.41	.40	.49	.47	.46
TEMI-PM PV	.49	.51	.49	.53	.55	.53
TEMI-PM ASC	.40	.40	.38	.46	.46	.44
TEMI-PM Total	.49	.44	.46	.56	.50	.52

© 2008 The University of Texas System/Texas Education Agency

Table 47. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-PM Scores and SAT-10 Primary-1 MPS Scores (Spring 2007 Administration)

	R with Primary-1 MPS					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.48	.55	.56	.54	.61	.62
TEMI-PM NS	.44	.53	.51	.48	.57	.55
TEMI-PM PV	.43	.49	.49	.47	.53	.53
TEMI-PM ASC	.53	.59	.60	.57	.63	.64
TEMI-PM Total	.56	.63	.64	.60	.67	.68

Table 48. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-PM Scores and SAT-10 Primary-1 MP Scores (Spring 2007 Administration)

	R with Primary-1 MP					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.39	.35	.40	.42	.38	.43
TEMI-PM NS	.38	.47	.44	.41	.51	.48
TEMI-PM PV	.42	.41	.45	.45	.44	.49
TEMI-PM ASC	.49	.53	.54	.53	.57	.58
TEMI-PM Total	.50	.51	.54	.54	.55	.58

Table 49. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-PM Scores and SAT-10 Primary-1 Total Scores (Spring 2007 Administration)

	R with Primary-1 Total					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.48	.55	.56	.53	.61	.61
TEMI-PM NS	.40	.47	.46	.43	.51	.50
TEMI-PM PV	.39	.37	.40	.42	.40	.43
TEMI-PM ASC	.36	.47	.44	.39	.51	.48
TEMI-PM Total	.54	.61	.61	.58	.65	.65

Table 50. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 Form C TEMI-PM Scores and SAT-10 Primary-1 Scores – Spring 2008 Administration

	R with Primary-1 Uncorrected			R with Primary-1 Corrected		
	MPS	MP	Total	MPS	MP	Total
TEMI-PM MC	.45	.46	.50	.51	.49	.55
TEMI-PM NS	.48	.51	.54	.54	.54	.59
TEMI-PM PV	.42	.46	.48	.48	.49	.53
TEMI-PM ASC	.50	.50	.54	.56	.53	.59
TEMI-PM Total	.56	.57	.62	.62	.60	.67

© 2008 The University of Texas System/Texas Education Agency

Table 51. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-O Scores and SAT-10 Primary-1 Scores—Spring 2007 Administration

SAT-10 Scores	TEMI-O Score Uncorrected			TEMI-O Score Corrected		
	MPS	C	Total	MPS	C	Total
MPS	.70	.35	.60	.76	.40	.66
MP	.59	.34	.64	.62	.37	.67
Total	.68	.38	.61	.70	.40	.64

Table 52. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 Form C TEMI-O Scores and SAT-10 Primary-1 Scores—Spring 2008 Administration

	TEMI-O Score Uncorrected			TEMI-O Score Corrected		
Score	MPS	C	Total	MPS	C	Total
MPS	71	57	73	.77	.63	.78
MP	56	62	69	.59	.65	.71
Total	69	64	77	.74	.69	.81

Table 53. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 Form C TEMI-PM Scores and SAT-10 Primary-2 Scores—Spring 2008 Administration

	R with Primary-1 Uncorrected			R with Primary-1 Corrected		
	MPS	MP	Total	MPS	MP	Total
TEMI-PM MC	.49	.50	.54	.56	.52	.59
TEMI-PM NS	.39	.50	.48	.46	.52	.53
TEMI-PM PV	.42	.46	.47	.49	.48	.51
TEMI-PM ASC	.47	.55	.55	.54	.57	.60
TEMI-PM Total	.53	.62	.63	.60	.64	.68

Table 54. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-O Scores and SAT-10 Primary-2 Scores (Spring 2008 Administration)

	R with SAT-10 Scores					
	Uncorrected			Corrected		
SAT-10 Scores	TEMI-O MPS	TEMI-O C	TEMI-O Total	TEMI-O MPS	TEMI-O C	Total
SAT-10 Primary-2 Math Problem Solving	.76	.66	.78	.82	.73	.83
SAT-10 Primary-2 Math Procedures	.63	.72	.76	.65	.73	.78
SAT-10 Primary-2 Math Total	.77	.74	.84	.80	.78	.88

© 2008 The University of Texas System/Texas Education Agency

Table 55. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores (Fall 2006 Administration) and SAT-10 Primary-1 Math Problem Solving Scores (Fall 2006 Administration)

	R with SAT-10 Math Problem Solving Scores					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.52	.38	.45	.65	.50	.58
TEMI-PM NS	.47	.40	.48	.51	.43	.52
TEMI-PM PV	.63	.66	.63	.67	.70	.67
TEMI-PM ASC	.63	.67	.58	.67	.71	.62
TEMI-PM Total	.68	.66	.62	.72	.70	.66

Table 56. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores (Winter 2007 Administration) and SAT-10 Primary-1 Math Problem Solving Scores (Fall 2006 Administration)

	R with SAT-10 Math Problem Solving Scores					
	Uncorrected			Corrected		
Score	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.47	.43	.66	.60	.56	.78
TEMI-PM NS	.45	.54	.50	.49	.58	.54
TEMI-PM PV	.54	.48	.57	.58	.52	.61
TEMI-PM ASC	.50	.60	.51	.54	.64	.55
TEMI-PM Total	.57	.57	.62	.61	.61	.66

Table 57. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores and SAT-10 Primary-1 Math Problem Solving Scores (Spring 2007 Administration)

	R with SAT-10 Math Problem Solving Scores					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.43	.36	.42	.48	.41	.47
TEMI-PM NS	.54	.49	.52	.59	.54	.57
TEMI-PM PV	.53	.41	.48	.58	.46	.53
TEMI-PM ASC	.56	.56	.55	.61	.61	.61
TEMI-PM Total	.61	.56	.57	.66	.61	.62

Table 58. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores (Fall 2006 Administration) and SAT-10 Primary-1 Math Procedures Scores (Fall 2006 Administration)

	R with SAT-10 Math Procedures Scores					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.45	.34	.41	.55	.43	.51
TEMI-PM NS	.49	.57	.40	.53	.61	.43
TEMI-PM PV	.55	.58	.54	.59	.61	.58
TEMI-PM ASC	.61	.65	.57	.65	.69	.61
TEMI-PM Total	.62	.61	.57	.66	.65	.61

© 2008 The University of Texas System/Texas Education Agency

Table 59. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores (Winter 2007 Administration) and SAT-10 Primary-1 Math Procedures Scores (Fall 2006 Administration)

	R with SAT-10 Math Procedures Scores					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.41	.38	.43	.51	.47	.53
TEMI-PM NS	.43	.53	.48	.47	.57	.52
TEMI-PM PV	.48	.43	.51	.52	.47	.55
TEMI-PM ASC	.46	.57	.50	.50	.61	.61
TEMI-PM Total	.54	.57	.57	.58	.61	.61

Table 60. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores and SAT-10 Primary-1 Math Procedures Scores (Spring 2007 Administration)

	R with SAT-10 Math Procedures Scores					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.40	.37	.42	.42	.39	.44
TEMI-PM NS	.49	.47	.49	.51	.49	.51
TEMI-PM PV	.49	.36	.42	.51	.38	.44
TEMI-PM ASC	.53	.51	.51	.55	.53	.53
TEMI-PM Total	.56	.51	.51	.58	.53	.53

Table 61. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores (Fall 2006 Administration) and SAT-10 Primary-1 Total Scores (Fall 2006 Administration)

	R with SAT-10 Total Scores					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.52	.38	.45	.65	.50	.58
TEMI-PM NS	.47	.40	.48	.51	.43	.52
TEMI-PM PV	.63	.66	.63	.67	.70	.67
TEMI-PM ASC	.63	.67	.58	.67	.71	.62
TEMI-PM Total	.68	.66	.62	.72	.70	.66

Table 62. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores (Winter 2007 Administration) and SAT-10 Primary-1 Total Scores (Fall 2006 Administration)

	R with SAT-10 Total Scores					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.47	.43	.66	.60	.56	.78
TEMI-PM NS	.45	.54	.50	.49	.58	.54
TEMI-PM PV	.54	.48	.57	.58	.52	.61
TEMI-PM ASC	.50	.60	.51	.54	.64	.55
TEMI-PM Total	.57	.57	.62	.61	.61	.66

© 2008 The University of Texas System/Texas Education Agency

Table 63. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores and SAT-10 Primary-2 Total Scores (Spring 2007 Administration)

	R with SAT-10 Total Scores					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.43	.36	.42	.48	.41	.47
TEMI-PM NS	.54	.49	.52	.59	.54	.57
TEMI-PM PV	.53	.41	.48	.58	.46	.53
TEMI-PM ASC	.56	.56	.55	.61	.61	.61
TEMI-PM Total	.61	.56	.57	.66	.61	.62

Table 64. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 Form C TEMI-PM Scores and SAT-10 Primary-2 Scores—Spring 2008 Administration

	R with Primary-2 Uncorrected			R with Primary-2 Corrected		
	MPS	MP	Total	MPS	MP	Total
TEMI-PM MC	.48	.44	.50	.55	.50	.55
TEMI-PM NS	.46	.51	.52	.52	.48	.57
TEMI-PM PV	.38	.39	.41	.44	.41	.45
TEMI-PM ASC	.27	.31	.32	.32	.32	.36
TEMI-PM Total	.56	.58	.61	.63	.60	.66

Table 65. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-O Scores and SAT-10 Primary-1 Scores (Fall 2007 Administration)

	R with SAT-10 Scores					
	Uncorrected			Corrected		
SAT-10 Scores	TEMI-O MPS	TEMI-O C	TEMI-O Total	TEMI-O MPS	TEMI-O C	TEMI-O Total
SAT-10 Primary-1 Math Problem Solving	.74	.64	.77	.78	.67	.79
SAT-10 Primary-1 Math Procedures	.62	.69	.74	.66	.71	.76
SAT-10 Primary-1 Math Total	.74	.71	.81	.78	.73	.76

© 2008 The University of Texas System/Texas Education Agency

Table 66. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-O MPS Scores (Winter 2007 Administration) and SAT-10 Primary-1 Scores (Fall 2006 Administration)

	R with SAT-10 Scores			
	Uncorrected		Corrected	
SAT-10 Scores	TEMI-O MPS Form A	TEMI-O MPS Form B	TEMI-O MPS Form A	TEMI-O MPS Form B
SAT-10 Primary-1 Math Problem Solving	.62	.63	.66	.67
SAT-10 Primary-1 Math Procedures	.70	.67	.72	.70
SAT-10 Primary-1 Math Total	.72	.70	.75	.73

Table 67. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-O Scores and SAT-10 Primary-2 Scores (Spring 2007 Administration)

	R with SAT-10 Scores					
	Uncorrected			Corrected		
SAT-10 Scores	TEMI-O MPS Form A	TEMI-O C Form B	TEMI-O Total	TEMI-O MPS Form A	TEMI-O C Form B	TEMI-O Total
SAT-10 Primary-2 Math Problem Solving	.61	.69	.71	.68	.75	.77
SAT-10 Primary-2 Math Procedures	.78	.72	.80	.80	.74	.82
SAT-10 Primary-2 Math Total	.77	.77	.83	.81	.81	.86

Table 68. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 Form C TEMI-O Scores and SAT-10 Primary-2 Scores (Spring 2008 Administration)

	R with SAT-10 Scores					
	Uncorrected			Corrected		
SAT-10 Scores	TEMI-O MPS	TEMI-O C	TEMI-O Total	TEMI-O MPS	TEMI-O C	TEMI-O Total
SAT-10 Primary-1 Math Problem Solving	.76	.60	.76	.79	.67	.81
SAT-10 Primary-1 Math Procedures	.64	.64	.74	.66	.66	.76
SAT-10 Primary-1 Math Total	.76	.67	.81	.80	.71	.84

© 2008 The University of Texas System/Texas Education Agency

Table 69. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Kindergarten TEMI-PM Scores (Fall 2006 Administration) and ITBS Scores (Fall 2006 Administration)

	R with ITBS Mathematics		
TEMI-PM Scores	Form A	Form B	Form C
TEMI-PM MC	.56	.50	.53
TEMI-PM NI	.55	.50	.51
TEMI-PM NS	.51	.47	.47
TEMI-PM QR	.49	.39	.43
TEMI-PM Total	.62	.57	.59

Table 70. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-PM Scores (Fall 2006 Administration) and ITBS Scores (Fall 2006 Administration)

	R with ITBS Mathematics		
TEMI-PM Scores	Form A	Form B	Form C
TEMI-PM MC	.45	.50	.39
TEMI-PM NI	.58	.53	.47
TEMI-PM NS	.59	.54	.49
TEMI-PM QR	.51	.62	.54
TEMI-PM Total	.62	.62	.55

Table 71. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores (Fall 2006 Administration) and ITBS Math Concepts Scores (Fall 2006 Administration)

	R with ITBS Mathematics		
TEMI-PM Scores	Form A	Form B	Form C
TEMI-PM MC	.38	.31	.37
TEMI-PM NI	.30	.35	.23
TEMI-PM NS	.49	.48	.44
TEMI-PM QR	.38	.40	.37
TEMI-PM Total	.44	.44	.42

Table 72. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores (Fall 2006 Administration) and ITBS Math Problems Scores (Fall 2006 Administration)

	R with ITBS Mathematics		
TEMI-PM Scores	Form A	Form B	Form C
TEMI-PM MC	.32	.28	.24
TEMI-PM NI	.28	.30	NS
TEMI-PM NS	.54	.53	.52
TEMI-PM QR	.33	.35	.33
TEMI-PM Total	.41	.41	.35

© 2008 The University of Texas System/Texas Education Agency

Table 73. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores (Fall 2006 Administration) and ITBS Math Computation Scores (Fall 2006 Administration)

	R with ITBS Mathematics		
TEMI-PM Scores	Form A	Form B	Form C
TEMI-PM MC	.37	.31	.32
TEMI-PM NI	.27	.34	.28
TEMI-PM NS	.33	.31	.41
TEMI-PM QR	.52	.50	.49
TEMI-PM Total	.45	.44	.45

Table 74. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores (Fall 2006 Administration) and ITBS Math Total Scores (Fall 2006 Administration)

	R with ITBS Mathematics		
TEMI-PM Scores	Form A	Form B	Form C
TEMI-PM MC	.44	.38	.38
TEMI-PM NI	.35	.40	.28
TEMI-PM NS	.58	.56	.57
TEMI-PM QR	.49	.50	.48
TEMI-PM Total	.54	.553	.50

Table 75. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Kindergarten TEMI-O Scores (Fall 2006 Administration) and ITBS Scores (Fall 2006 Administration)

	R with ITBS Mathematics
TEMI-O Scores	Form A
TEMI-O MPS	.50

Table 76. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-O Scores (Fall 2006 Administration) and ITBS Scores (Fall 2006 Administration)

	R with ITBS Mathematics
TEMI-O Scores	Form A
TEMI-O MPS	.65
TEMI-O C	.66
TEMI-O Total	.72

Table 77. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-O Scores (Fall 2006 Administration) and ITBS Math Concepts Scores (Fall 2006 Administration)

	R with ITBS Math Concepts
TEMI-O Scores	Form A
TEMI-O MPS	.61
TEMI-O C	.50
TEMI-O Total	.65

© 2008 The University of Texas System/Texas Education Agency

Table 78. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-O Scores (Fall 2006 Administration) and ITBS Math Problems Scores (Fall 2006 Administration)

	R with ITBS Math Concepts
TEMI-O Scores	Form A
TEMI-O MPS	.62
TEMI-O C	.32
TEMI-O Total	.56

Table 79. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-O Scores (Fall 2006 Administration) and ITBS Math Computation Scores (Fall 2006 Administration)

	R with ITBS Math Computation
TEMI-O Scores	Form A
TEMI-O MPS	.27
TEMI-O C	.37
TEMI-O Total	.38

Table 80. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-O Scores (Fall 2006 Administration) and ITBS Math Total Scores (Fall 2006 Administration)

	R with ITBS Math Total
TEMI-O Scores	Form A
TEMI-O MPS	.67
TEMI-O C	.49
TEMI-O Total	.68

Table 81. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-PM Scores and Math Teacher Ratings—Fall 2007 Administration

TEMI-PM Scores	R with Teacher Ratings
TEMI-PM MC	.44
TEMI-PM NS	.45
TEMI-PM PV	.43
TEMI-PM ASC	.39
TEMI-PM Total	.51

Table 82. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores and Math Teacher Ratings – Fall 2007 Administration

TEMI-PM Scores	R with Teacher Ratings
TEMI-PM MC	.42
TEMI-PM NS	.48
TEMI-PM PV	.47
TEMI-PM ASC	.49
TEMI-PM Total	.55

© 2008 The University of Texas System/Texas Education Agency

Table 83. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Kindergarten TEMI-PM Scores (Fall 2006 Administration) and Teacher Math Ratings (Fall 2006)

	R with ITBS Mathematics		
TEMI-PM Scores	Form A	Form B	Form C
TEMI-PM MC	.41	.43	.45
TEMI-PM NI	.49	.53	.56
TEMI-PM NS	.44	.46	.44
TEMI-PM QR	.32	.36	.36
TEMI-PM Total	.48	.50	.48

Table 84. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-PM Scores (Fall 2006 Administration) and Teacher Math Ratings (Fall 2006)

	R with ITBS Mathematics		
TEMI-PM Scores	Form A	Form B	Form C
TEMI-PM MC	.37	NS	.32
TEMI-PM NS	.33	.35	.36
TEMI-PM PV	.27	.33	.22
TEMI-PM ASC	.29	.27	.22
TEMI-PM Total	.39	.31	.39

Table 85. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-O Scores and Math Teacher Ratings—Fall 2007 Administration

TEMI-PM Scores	R with Teacher Ratings
TEMI-O MPS	.51
TEMI-O C	.48
TEMI-O Total	.57

Table 86. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-O Scores and Math Teacher Ratings—Fall 2007 Administration

TEMI-PM Scores	R with Teacher Ratings
TEMI-O MPS	.58
TEMI-O C	.52
TEMI-O Total	.61

Table 87. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores (Fall 2006 Administration) and Teacher LDDI Math Ratings (Fall 2006)

	R with ITBS Mathematics		
TEMI-PM Scores	Form A	Form B	Form C
TEMI-PM MC	.41	.40	.39
TEMI-PM NS	.37	.29	NS
TEMI-PM PV	.32	NS	NS
TEMI-PM ASC	.44	.42	.45
TEMI-PM Total	.50	.44	.44

© 2008 The University of Texas System/Texas Education Agency

Table 88. Concurrent Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-O Scores (Fall 2006 Administration) and LDDI Math Scores (Fall 2006 Administration)

	R with ITBS Math Total
TEMI-O Scores	Form A
TEMI-O MPS	.68
TEMI-O C	.51
TEMI-O Total	.63

The coefficients reported above for the concurrent criterion-prediction validity of the TEMI-PM and TEMI-O provide support for the validity of the test scores reported for the two measures.

Predictive Criterion-Prediction Validity

To examine the predictive aspect of criterion-prediction validity of the TEMI-O and TEMI-PM, we conducted two types of studies. First, we correlated the fall and winter administrations of the TEMI-PM and TEMI-O with the spring scores made by the students on the SAT-10. Second, we examined the Area Under the Curve (AUC) base on Receiver Operator Curve analyses, again using the SAT-10 as the criterion measure. Each of these analyses is discussed here.

Correlations

As noted previously, during the fall of the 2006–2007 and 2007–2008 academic years, we administered TEMI-PM and the TEMI-O. In the spring of those academic years, we administered the mathematics subtests of the Stanford Achievement Test – 10th edition (SAT-10). We correlated the results of these administrations, and the resulting coefficients are depicted below. Note again that when we report correlations with the SAT-10, we provide two coefficients, one of which reporting the relationship between the measures after accounting for restricted range of the students who participated in the study.

Table 89. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Kindergarten TEMI-PM Scores (Fall 2006 Administration) and SAT-10 SESAT-2 Scores (Spring 2007 Administration)

	R with SESAT-2					
	Uncorrected			Corrected		
Score	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.58	.56	.55	.64	.62	.61
TEMI-PM NI	.51	.50	.60	.57	.56	.66
TEMI-PM NS	.61	.62	.62	.67	.68	.68
TEMI-PM QR	.48	.47	.53	.54	.53	.59
TEMI-PM Total	.65	.63	.66	.71	.69	.71

© 2008 The University of Texas System/Texas Education Agency

Table 90. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Kindergarten TEMI-PM Scores (Winter 2007 Administration) and SAT-10 SESAT-2 Scores (Spring 2007 Administration)

	R with SESAT-2					
	Uncorrected			Corrected		
Score	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.54	.52	.50	.60	.58	.56
TEMI-PM NI	.51	.46	.46	.57	.52	.52
TEMI-PM NS	.49	.57	.56	.55	.63	.62
TEMI-PM QR	.39	.44	.41	.44	.50	.46
TEMI-PM Total	.57	.60	.56	.63	.66	.62

Table 91. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Kindergarten TEMI-PM Scores (Fall 2007 Administration) and SAT-10 SESAT-2 Scores (Spring 2008 Administration)

	R with SESAT-2	
	Uncorrected	Corrected
Score	Form A	Form A
Fall - TEMI-PM MC	.62	.72
Fall - TEMI-PM NI	.58	.68
Fall - TEMI-PM NS	.58	.68
Fall - TEMI-PM QR	.54	.64
Fall - TEMI-PM Total	.67	.76

Table 92. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Kindergarten TEMI-PM Scores (Winter 2008 Administration) and SAT-10 SESAT -2 Scores (Spring 2008 Administration)

	R with SESAT-2	
	Uncorrected	Corrected
Score	Form A	Form A
TEMI-PM MC	.53	.63
TEMI-PM NI	.56	.66
TEMI-PM NS	.52	.62
TEMI-PM QR	.43	.53
TEMI-PM Total	.60	.70

Table 93. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Kindergarten TEMI-O Scores (Fall 2006 Administration) and SAT-10 SESAT-2 Scores (Spring 2007 Administration)

	R with SESAT-2	
	Uncorrected	Corrected
Score	Form A	Form A
TEMI-O MPS	.67	.72

© 2008 The University of Texas System/Texas Education Agency

Table 94. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Kindergarten TEMI-O Scores (Winter 2007 Administration) and SAT-10 SESAT-2 Scores (Spring 2007 Administration)

	R with SESAT-2	
	Uncorrected	Corrected
Score	Form A	Form A
TEMI-O MPS	.73	.78

Table 95. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Kindergarten TEMI-O Scores (Fall 2007 Administration) and SAT-10 SESAT-2 Scores (Spring 2008 Administration)

	R with SESAT-2	
	Uncorrected	Corrected
Score	Form A	Form A
TEMI-O MPS	.63	.72

Table 96. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Kindergarten TEMI-O Scores (Winter 2008 Administration) and SAT-10 SESAT-2 Scores (Spring 2008 Administration)

	R with SESAT-2	
	Uncorrected	Corrected
Score	Form A	Form A
TEMI-O MPS	.68	.77

Table 97. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 Form A TEMI-O Scores (Fall 2006 Administration) and SAT-10 Primary-1 Math Problem Solving Scores (Spring 2007 Administration)

	R with SESAT-2	
	Uncorrected	Corrected
Score	Form A	Form A
TEMI-O MPS	.73	.78

Table 98. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-PM Scores (Fall 2006 Administration) and SAT-10 Primary-1 Math Problem Solving Scores (Spring 2007 Administration)

	R with Primary-1					
	Uncorrected			Corrected		
Score	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.54	.52	.50	.60	.58	.56
TEMI-PM NS	.51	.46	.46	.57	.52	.52
TEMI-PM PV	.49	.57	.56	.55	.63	.62
TEMI-PM ASC	.39	.44	.41	.44	.50	.46
TEMI-PM Total	.57	.60	.56	.63	.66	.62

© 2008 The University of Texas System/Texas Education Agency

Table 99. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-PM Scores (Fall 2006 Administration) and SAT-10 Primary-1 Math Problem Solving Scores (Spring 2007 Administration)

	R with Primary-1					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.49	.55	.63	.54	.61	.68
TEMI-PM NS	.57	.53	.55	.59	.58	.56
TEMI-PM PV	.62	.53	.65	.65	.56	.68
TEMI-PM ASC	.56	.55	.53	.59	.58	.56
TEMI-PM Total	.68	.56	.68	.71	.59	.71

Table 100. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-PM Scores (Fall 2006 Administration) and SAT-10 Primary-1 Math Procedures Scores (Spring 2007 Administration)

	R with Primary-1					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.53	.56	.61	.59	.62	.67
TEMI-PM NS	.50	.48	.53	.56	.54	.59
TEMI-PM PV	.59	.52	.61	.65	.58	.67
TEMI-PM ASC	.60	.55	.55	.66	.61	.61
TEMI-PM Total	.68	.59	.66	.74	.65	.72

Table 101. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-PM Scores (Fall 2006 Administration) and SAT-10 Primary-1 Total Scores (Spring 2007 Administration)

	R with Primary-1					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.37	.47	.53	.60	.58	.56
TEMI-PM NS	.61	.58	.60	.66	.63	.65
TEMI-PM PV	.52	.54	.54	.57	.59	.59
TEMI-PM ASC	.63	.50	.63	.68	.55	.68
TEMI-PM Total	.70	.63	.70	.75	.68	.75

Table 102. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-PM Scores (Winter 2007 Administration) and SAT-10 Primary-1 Math Problem Solving Scores (Spring 2007 Administration)

	R with Primary-1					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.74	.65	.72	.47	.38	.43
TEMI-PM NS	.49	.55	.47	.52	.58	.50
TEMI-PM PV	.60	.55	.59	.63	.58	.62
TEMI-PM ASC	.54	.56	.51	.57	.59	.54
TEMI-PM Total	.60	.57	.56	.63	.60	.59

© 2008 The University of Texas System/Texas Education Agency

Table 103. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-PM Scores (Winter 2007 Administration) and SAT-10 Primary-1 Math Procedures Scores (Spring 2007 Administration)

	R with Primary-1					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.41	.33	.33	.47	.38	.38
TEMI-PM NS	.49	.48	.40	.55	.54	.46
TEMI-PM PV	.55	.48	.54	.61	.54	.60
TEMI-PM ASC	.56	.56	.46	.62	.62	.52
TEMI-PM Total	.58	.52	.49	.64	.58	.55

Table 104. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-PM Scores (Winter 2007 Administration) and SAT-10 Primary-1 Total Scores (Spring 2007 Administration)

	R with Primary-1					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.41	.36	.41	.46	.41	.46
TEMI-PM NS	.50	.52	.42	.55	.57	.47
TEMI-PM PV	.56	.53	.56	.61	.58	.61
TEMI-PM ASC	.56	.53	.46	.61	.58	.51
TEMI-PM Total	.60	.56	.55	.65	.61	.61

Table 105. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 Form A TEMI-PM Scores (Fall 2007 Administration) and SAT-10 Primary-1 Math Problem Solving Scores (Spring 2008 Administration)

	R with Primary-1	
	Uncorrected	Corrected
TEMI-PM Scores	Form B	Form B
TEMI-PM MC	.49	.56
TEMI-PM NS	.51	.58
TEMI-PM PV	.41	.47
TEMI-PM ASC	.50	.57
TEMI-PM Total	.57	.64

Table 106. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 Form A TEMI-PM Scores (Winter 2008 Administration) and SAT-10 Primary-1 Math Procedures Scores (Spring 2008 Administration)

	R with Primary-1	
	Uncorrected	Corrected
TEMI-PM Scores	Form B	Form B
TEMI-PM MC	.37	.39
TEMI-PM NS	.44	.46
TEMI-PM PV	.32	.34
TEMI-PM ASC	.46	.48
TEMI-PM Total	.46	.48

© 2008 The University of Texas System/Texas Education Agency

Table 107. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 Form A TEMI-PM Scores (Winter 2008 Administration) and SAT-10 Primary-1 Total Scores (Spring 2008 Administration)

	R with Primary-1	
	Uncorrected	Corrected
TEMI-PM Scores	Form B	Form B
TEMI-PM MC	.47	.52
TEMI-PM NS	.52	.57
TEMI-PM PV	.41	.45
TEMI-PM ASC	.53	.58
TEMI-PM Total	.57	.62

Table 108. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 Form A TEMI-PM Scores (Winter 2007 Administration) and SAT-10 Primary-1 Math Procedures Scores (Spring 2008 Administration)

	R with Primary-1	
	Uncorrected	Corrected
TEMI-PM Scores	Form B	Form B
TEMI-PM MC	.43	.45
TEMI-PM NS	.48	.50
TEMI-PM PV	.42	.44
TEMI-PM ASC	.57	.59
TEMI-PM Total	.56	.58

Table 109. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 Form A TEMI-PM Scores (Winter 2008 Administration) and SAT-10 Primary-1 Total Scores (Spring 2008 Administration)

	R with Primary-1	
	Uncorrected	Corrected
TEMI-PM Scores	Form B	Form B
TEMI-PM MC	.48	.53
TEMI-PM NS	.52	.57
TEMI-PM PV	.49	.54
TEMI-PM ASC	.61	.66
TEMI-PM Total	.61	.66

© 2008 The University of Texas System/Texas Education Agency

Table 110. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-O Scores (Fall 2007 Administration) and SAT-10 Primary-1 Scores (Spring 2008 Administration)

	R with SAT-10 Scores					
	Uncorrected			Corrected		
SAT-10 Scores	TEMI-O MPS	TEMI-O C	TEMI-O Total	TEMI-O MPS	TEMI-O C	TEMI-O Total
SAT-10 Primary-1 Math Problem Solving	.57	.59	.66	.64	.66	.72
SAT-10 Primary-1 Math Procedures	.37	.53	.52	.39	.55	.44
SAT-10 Primary-1 Math Total	.52	.62	.66	.57	.67	.71

Table 111. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 1 TEMI-O Scores (Winter 2007 Administration) and SAT-10 Primary-1 Scores (Spring 2008 Administration)

	R with SAT-10 Scores					
	Uncorrected			Corrected		
SAT-10 Scores	TEMI-O MPS	TEMI-O C	TEMI-O Total	TEMI-O MPS	TEMI-O C	TEMI-O Total
SAT-10 Primary-1 Math Problem Solving	.65	.58	.70	.71	.65	.76
SAT-10 Primary-1 Math Procedures	.47	.52	.56	.49	.54	.58
SAT-10 Primary-1 Math Total	.62	.60	.69	.67	.65	.73

Table 112. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores (Fall 2006 Administration) and SAT-10 Primary-2 Math Problem Solving Scores (Spring 2007 Administration)

	R with Primary-2					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.44	.26	.36	.50	.31	.43
TEMI-PM NS	.41	.52	.39	.48	.60	.45
TEMI-PM PV	.59	.60	.61	.67	.67	.68
TEMI-PM ASC	.56	.58	.51	.63	.65	.58
TEMI-PM Total	.58	.54	.53	.65	.61	.60

© 2008 The University of Texas System/Texas Education Agency

Table 113. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores (Fall 2006 Administration) and SAT-10 Primary-2 Math Procedures Scores (Spring 2007 Administration)

	R with Primary-2					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.35	.25	.30	.37	.26	.32
TEMI-PM NS	.36	.43	.29	.38	.45	.30
TEMI-PM PV	.48	.51	.41	.50	.53	.43
TEMI-PM ASC	.54	.55	.43	.56	.57	.45
TEMI-PM Total	.51	.49	.42	.53	.51	.44

Table 114. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores (Fall 2006 Administration) and SAT-10 Primary-2 Total Scores (Spring 2007 Administration)

	R with Primary-2					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.43	.28	.36	.48	.32	.41
TEMI-PM NS	.42	.53	.38	.47	.58	.43
TEMI-PM PV	.58	.60	.56	.63	.65	.61
TEMI-PM ASC	.60	.62	.51	.65	.67	.56
TEMI-PM Total	.60	.57	.52	.65	.62	.57

Table 115. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores (Winter 2007 Administration) and SAT-10 Primary-2 Math Problem Solving Scores (Spring 2007 Administration)

	R with Primary-2					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.38	.33	.38	.50	.44	.50
TEMI-PM NS	.38	.38	.42	.50	.50	.54
TEMI-PM PV	.48	.41	.49	.61	.53	.62
TEMI-PM ASC	.45	.45	.49	.57	.57	.62
TEMI-PM Total	.51	.47	.53	.64	.59	.66

Table 116. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores (Winter 2007 Administration) and SAT-10 Primary-1 Math Procedures Scores (Spring 2007 Administration)

	R with Primary-2					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.35	.33	.37	.44	.41	.46
TEMI-PM NS	.44	.43	.48	.54	.53	.58
TEMI-PM PV	.43	.38	.44	.53	.47	.54
TEMI-PM ASC	.53	.55	.59	.63	.65	.69
TEMI-PM Total	.52	.51	.56	.62	.61	.66

© 2008 The University of Texas System/Texas Education Agency

Table 117. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-PM Scores (Winter 2007 Administration) and SAT-10 Primary-1 Total Scores (Spring 2007 Administration)

	R with Primary-2					
	Uncorrected			Corrected		
TEMI-PM Scores	Form A	Form B	Form C	Form A	Form B	Form C
TEMI-PM MC	.41	.33	.41	.54	.37	.54
TEMI-PM NS	.44	.42	.47	.57	.55	.60
TEMI-PM PV	.52	.40	.51	.65	.53	.64
TEMI-PM ASC	.49	.50	.54	.62	.63	.67
TEMI-PM Total	.54	.48	.56	.67	.61	.69

Table 118. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 Form A TEMI-PM Scores (Fall 2007 Administration) and SAT-10 Primary-2 Math Problem Solving Scores (Spring 2008 Administration)

	R with Primary-2	
TEMI-PM Scores	Uncorrected Form A	Corrected Form A
Fall - TEMI-PM MC	.51	.59
Fall - TEMI-PM NS	.47	.54
Fall - TEMI-PM PV	.64	.71
Fall - TEMI-PM ASC	.52	.60
Fall - TEMI-PM Total	.61	.68

Table 119. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 Form A TEMI-PM Scores (Fall 2007 Administration) and SAT-10 Primary-2 Math Procedures Scores (Spring 2008 Administration)

	R with Primary-2	
TEMI-PM Scores	Uncorrected Form A	Corrected Form A
TEMI-PM MC	.51	.52
TEMI-PM NS	.51	.52
TEMI-PM PV	.59	.61
TEMI-PM ASC	.59	.61
TEMI-PM Total	.65	.67

Table 120. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 Form A TEMI-PM Scores (Fall 2007 Administration) and SAT-10 Primary-2 Total Scores (Spring 2008 Administration)

	R with Primary-2	
TEMI-PM Scores	Uncorrected Form A	Corrected Form A
TEMI-PM MC	.54	.58
TEMI-PM NS	.49	.53
TEMI-PM PV	.67	.71
TEMI-PM ASC	.58	.62
TEMI-PM Total	.67	.71

© 2008 The University of Texas System/Texas Education Agency

Table 121. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 Form B TEMI-PM Scores (Winter 2008 Administration) and SAT-10 Primary-2 Math Problem Solving Scores (Spring 2008 Administration)

	R with Primary-2	
TEMI-PM Scores	Uncorrected Form B	Corrected Form B
TEMI-PM MC	.49	.56
TEMI-PM NS	.41	.48
TEMI-PM PV	.40	.47
TEMI-PM ASC	.53	.60
TEMI-PM Total	.58	.65

Table 122. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 Form B TEMI-PM Scores (Winter 2008 Administration) and SAT-10 Primary-2 Math Procedures Scores (Spring 2008 Administration)

	R with Primary-2	
TEMI-PM Scores	Uncorrected Form B	Corrected Form B
TEMI-PM MC	.50	.52
TEMI-PM NS	.51	.53
TEMI-PM PV	.44	.45
TEMI-PM ASC	.62	.64
TEMI-PM Total	.65	.66

Table 123. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 Form B TEMI-PM Scores (Winter 2008 Administration) and SAT-10 Primary-2 Total Scores (Spring 2008 Administration)

	R with Primary-2	
TEMI-PM Scores	Uncorrected Form B	Corrected Form B
TEMI-PM MC	.53	.57
TEMI-PM NS	.48	.52
TEMI-PM PV	.44	.48
TEMI-PM ASC	.61	.65
TEMI-PM Total	.65	.69

© 2008 The University of Texas System/Texas Education Agency

Table 124. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-O Scores (Fall 2006 Administration) and SAT-10 Primary-2 Scores (Spring 2007 Administration)

	R with SAT-10 Scores					
	Uncorrected			Corrected		
SAT-10 Scores	TEMI-O MPS Form A	TEMI-O C Form A	TEMI-O Total	TEMI-O MPS Form A	TEMI-O C Form A	TEMI-O Total
SAT-10 Primary-1 Math Problem Solving	.69	.63	.73	.75	.69	.79
SAT-10 Primary-1 Math Procedures	.54	.57	.62	.56	.59	.64
SAT-10 Primary-1 Total	.68	.66	.74	.73	.71	.79

Table 125. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 TEMI-O Scores (Winter 2007 Administration) and SAT-10 Primary-2 Scores (Spring 2007 Administration)

	R with SAT-10 Scores			
	Uncorrected		Corrected	
SAT-10 Scores	TEMI-O MPS Form A	TEMI-O MPS Form B	TEMI-O MPS Form A	TEMI-O MPS Form B
SAT-10 Primary-1 Math Problem Solving	.71	.67	.77	.74
SAT-10 Primary-1 Math Procedures	.56	.56	.58	.58
SAT-10 Primary-1 Math Total	.69	.66	.73	.70

Table 126. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 Form A TEMI-O Scores (Fall 2007 Administration) and SAT-10 Primary-2 Scores (Spring 2008 Administration)

	R with SAT-10 Scores					
	Uncorrected			Corrected		
SAT-10 Scores	TEMI-O MPS	TEMI-O MPS	TEMI-O Total	TEMI-O MPS	TEMI-O MPS	TEMI-O Total
SAT-10 Primary-1 Math Problem Solving	.67	.63	.73	.74	.70	.79
SAT-10 Primary-1 Math Procedures	.53	.63	.67	.55	.65	.69
SAT-10 Primary-1 Math Total	.66	.68	.76	.70	.72	.79

© 2008 The University of Texas System/Texas Education Agency

Table 127. Predictive Criterion-Prediction Validity Correlation Coefficients Depicting the Relationship Between Grade 2 Form B TEMI-O Scores (Winter 2008 Administration) and SAT-10 Primary-2 Scores (Spring 2008 Administration)

	R with SAT-10 Scores					
	Uncorrected			Corrected		
SAT-10 Scores	TEMI-O MPS	TEMI-O MPS	TEMI-O Total	TEMI-O MPS	TEMI-O MPS	TEMI-O Total
SAT-10 Primary-1 Math Problem Solving	.73	.70	.78	.79	.77	.83
SAT-10 Primary-1 Math Procedures	.61	.73	.75	.63	.74	.76
SAT-10 Primary-1 Math Total	.74	.76	.83	.78	.79	.86

The coefficients reported above provide support for the predictive criterion-prediction validity of the test scores reported for the TEMI-PM and TEMI-O.

Area Under the Curve

One way to examine the predictive validity of a measure is to examine how accurately the test separates the group being tested into those with and without the condition in question—in this case, mathematics difficulties. Accuracy, or prediction reliability, as it is also called, is measured by the area under the Receiver Operating Curve (ROC). Originally designed to estimate the accuracy with which radar could identify incoming planes in time of war, ROC analyses have been adopted in medicine to diagnose the likelihood that a person has a particular disease based on test results. Education has most recently borrowed the procedure to examine the prediction accuracy of screening tools to predict the presence of a reading problem, and we borrow it here to examine the prediction accuracy of the TEMI-PM and TEMI-O.

Dr. Steve Simon of Children's Mercy Hospitals and Clinics provides guidelines (www.childrens-mercy.org/stats/ask/roc.asp) for classifying the accuracy of a diagnostic test (an area of 1 represents a perfect test; an area of .5 represents a worthless test):

- 0.50 to 0.75 = fair
- 0.75 to 0.92 = good
- 0.92 to 0.97 = very good
- 0.97 to 1.00 = excellent

We conducted ROC analyses for the subtests and total scores for the TEMI-PM and TEMI-O, using the SAT-10 as our criterion measure (a score of below 90 on MPS in kindergarten and the Total Mathematics for grades 1 and 2 represents math difficulties). The results of the analysis are found below. As can be seen, most of the values indicate Good utility. Of particular interest are the total scores, because these scores are used for identification and progress monitoring purposes; and all of which demonstrate Good utility.

© 2008 The University of Texas System/Texas Education Agency

Table 128. Summary Table Including Utility and Area Under the Curve, Kindergarten

TEMI Score	Utility	Area	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
MC Fall	Fair	.74	.032	.677	.803
NID Fall	Fair	.74	.034	.676	.808
NS Fall	Fair	.74	.031	.676	.799
QR Fall	Good	.76	.033	.691	.821
Total PM Fall	Good	.80	.030	.739	.855
MPS O Fall	Good	.80	.030	.741	.857
MC Winter	Good	.76	.033	.698	.827
NID Winter	Good	.78	.031	.722	.843
NS Winter	Good	.76	.036	.684	.826
QR Winter	Fair	.71	.038	.634	.781
Total PM Winter	Good	.80	.034	.731	.863
MPS O Winter	Good	.86	.025	.807	.905
MC Spring	Fair	.72	.038	.648	.799
NID Spring	Fair	.73	.039	.653	.805
NS Spring	Good	.76	.033	.696	.826
QR Spring	Fair	.67	.040	.577	.734
Total PM Spring	Good	.76	.035	.691	.829
MPS O Spring	Good	.83	.028	.771	.882

© 2008 The University of Texas System/Texas Education Agency

Table 129. Summary Table Including Utility and Area Under the Curve, Grade 1

TEMI Score				95% Confidence Interval	
	Utility	Area	Std. Error	Lower Bound	Upper Bound
MC Fall	Good	.76	.029	.706	.820
NS Fall	Good	.76	.030	.695	.812
PV Fall	Fair	.67	.033	.612	.740
ASC Fall	Good	.76	.028	.705	.816
Total PM Fall	Good	.80	.027	.736	.844
MPS Fall	Good	.81	.026	.759	.862
C Fall	Good	.78	.028	.720	.831
Total O Fall	Good	.84	.024	.793	.888
MC Winter	Fair	.70	.032	.640	.767
NS Winter	Fair	.73	.031	.672	.793
PV Winter	Fair	.75	.031	.690	.812
ASC Winter	Good	.79	.028	.736	.845
Total PM Winter	Good	.78	.028	.728	.837
MPS Winter	Good	.85	.022	.810	.898
C Winter	Good	.80	.029	.730	.845
Total O Winter	Good	.85	.023	.803	.895
MC Spring	Fair	.71	.031	.648	.771
NS Spring	Fair	.72	.033	.654	.782
PV Spring	Fair	.69	.034	.619	.752
ASC Spring	Fair	.75	.031	.685	.805
Total PM Spring	Fair	.75	.031	.694	.814
MPS Spring	Good	.86	.023	.813	.901
C Spring	Good	.83	.026	.782	.883
Total O Spring	Good	.88	.022	.840	.925

© 2008 The University of Texas System/Texas Education Agency

Table 130. Summary Table Including Utility and Area Under the Curve, Grade 2

TEMI Score	Utility	Area	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
MC Fall	Good	.77	.031	.705	.826
NS Fall	Fair	.73	.032	.669	.794
PV Fall	Good	.82	.027	.770	.877
ASC Fall	Good	.80	.029	.742	.856
Total PM Fall	Good	.84	.027	.782	.887
MPS Fall	Good	.86	.023	.818	.908
C Fall	Good	.84	.025	.792	.889
Total O Fall	Good	.90	.019	.859	.935
MC Winter	Fair	.75	.032	.688	.813
NS Winter	Fair	.72	.033	.658	.789
PV Winter	Fair	.65	.039	.577	.730
ASC Winter	Good	.80	.028	.750	.859
Total PM Winter	Good	.80	.029	.747	.860
MPS Winter	Good	.86	.025	.812	.910
C Winter	Good	.87	.024	.820	.913
Total O Winter	Good	.90	.020	.860	.940
MC Spring	Good	.77	.030	.713	.833
NS Spring	Fair	.71	.034	.641	.775
PV Spring	Fair	.69	.037	.614	.758
ASC Spring	Fair	.73	.032	.671	.796
Total PM Spring	Good	.78	.029	.726	.841
MPS Spring	Good	.91	.019	.873	.946
C Spring	Good	.84	.026	.792	.893
Total O Spring	Good	.91	.018	.879	.949

Construct-Identification Validity

Anastasi and Urbina (1997) define construct-identification validity as "the extent to which the test may be said to measure a theoretical construct or trait" (p. 126). Test authors identify underlying traits of a test and then examine the extent to which their test reflects these traits.

We used Linn and Gronlund's (1995) three-step procedure to conduct our validity studies. First, we identified eight basic constructs that are associated with test performance on the TEMI-O and TEMI-PM. Second, we pair with each construct with a related hypothesis and question. Third, we examine each hypothesis using logical or empirical means.

1. Hypothesis: Math skills are developmental. Therefore, a test of mathematics achievement should produce scores that increase over the course of a year. Question: Do TEMI-PM and TEMI-O scores increase across the course of a year?
2. Hypothesis: Progress-monitoring tools should have variance that allows for low and high achievement. Therefore, floor and ceiling effects should not be present. Question: Are there floor and ceiling effects on the TEMI-PM and TEMI-O?
3. Hypothesis: Math skills are interrelated. Therefore, subtest scores should intercorrelate to a moderate to high degree. Question 3a: Do subtests intercorrelate with one another?

© 2008 The University of Texas System/Texas Education Agency

Question 3b: Do the scores load on a single factor, based on exploratory factor analytic research?

4. Hypothesis: Some students struggle in mathematics. Therefore, students who are known to be poor in mathematics should score lower than students who do well in mathematics. Question: Do the TEMI-PM and TEMI-O differentiate between low and average-to-high achievers?
5. Hypothesis: Mathematics is related to written language. Therefore, mathematics test scores should correlate to reading and writing test scores. Question 5a: Do the TEMI-PM and TEMI-O scores relate significantly to reading scores? Question 5b: Do the TEMI-PM and TEMI-O scores relate significantly to writing scores?
6. Hypothesis: Mathematics is one aspect of school achievement. Therefore, mathematics test scores should correlate to total scores on achievement tests. Question: Do the TEMI-PM and TEMI-O scores relate significantly to total achievement test scores?
7. Hypothesis: Mathematics tests are composed of items that are valid. Therefore, mathematics items within a subtest should intercorrelate with one another. Question: Do the TEMI-PM and TEMI-O items satisfy item validity criteria?
8. Hypothesis: Math skills are developmental. Therefore, a test of mathematics achievement should produce scores that increase over the course of a year. Question: Do TEMI-PM and TEMI-O scores increase across the course of a year?

Developmental Nature of Mathematics

Because mathematics is a development construct, that is, students as a rule improve their math skills as a result of instruction, math test scores should increase as students get older. When test scores are used to monitor progress as a result of intervention, it is important that scores improve as the academic year progresses. Chard et al. (2005) referred to differences across time as "change factors" and suggested that the differences be pronounced enough to reflect children's growth in skill attainment.

To test this hypothesis, we examined TEMI-PM and TEMI-O test scores in the fall, winter, and spring. Tables below provide mean scores at each point of the academic year. The fact that the average scores, especially the total scores, increase as they do provides support that the TEMI-PM and TEMI-O are sensitive enough to be used for progress-monitoring purposes, and also provide evidence for the construct-identification validity of the tests' scores.

Table 131. TEMI-PM Means and Standard Deviations by Season—Kindergarten

	Fall		Winter		Spring	
Test	Mean	SD	Mean	SD	Mean	SD
MC	16.51	13.491	29.27	14.856	38.31	15.133
NID	7.94	4.388	12.46	3.903	14.61	4.104
NS	6.87	6.724	13.42	7.616	18.44	8.163
QR	22.25	11.652	29.33	10.659	33.62	10.412
Total	53.57	31.316	84.48	30.997	104.31	32.477

© 2008 The University of Texas System/Texas Education Agency

Table 132. TEMI-PM Means and Standard Deviations by Season—Grade 1

	Fall		Winter		Spring	
Test	Mean	SD	Mean	SD	Mean	SD
MC	25.18	11.524	33.70	10.434	38.59	11.171
NS	11.86	6.299	18.27	7.210	20.31	7.371
PV	7.08	3.802	13.24	5.205	18.10	5.668
ASC	5.76	4.627	13.52	6.858	18.53	7.614
Total	49.88	22.191	78.75	25.358	95.24	27.163

Table 133. TEMI-PM Means and Standard Deviations by Season—Grade 2

	Fall		Winter		Spring	
Test	Mean	SD	Mean	SD	Mean	SD
MC	32.50	10.152	40.82	10.233	46.11	10.822
NS	12.25	4.156	18.14	5.680	19.44	6.573
PV	7.96	3.827	11.40	4.868	14.23	5.473
ASC	19.37	7.371	25.46	7.796	30.18	8.066
Total	72.07	20.537	95.83	23.208	109.93	25.336

Table 134. TEMI-O Means and Standard Deviations by Season—Kindergarten

	Fall		Winter		Spring	
Test	Mean	SD	Mean	SD	Mean	SD
MPS	20.00	6.561	25.01	5.624	26.23	5.101

Table 135. TEMI-O Means and Standard Deviations by Season—Grade 1

	Fall		Winter		Spring	
Test	Mean	SD	Mean	SD	Mean	SD
MPS	20.70	5.049	24.68	5.333	27.65	5.832
C	10.32	4.713	16.75	5.716	19.70	5.711
Total	30.70	8.809	41.03	10.203	46.99	10.846

Table 136. TEMI-O Means and Standard Deviations by Season—Grade 2

	Fall		Winter		Spring	
Test	Mean	SD	Mean	SD	Mean	SD
MPS	21.57	11.812	25.63	5.039	27.72	5.151
C	25.23	8.024	30.25	7.539	36.17	8.305
Total	46.30	11.812	55.15	11.990	63.48	12.617

Floor and Ceiling Effects

Hammill and Bryant (2005) discussed the importance of floor effects on measures used for diagnostic purposes. They referred to Bracken (1987) and Rathvon (2004), who offered that measures, for diagnostic purposes, should possess standard scores that range to at least 2 standard deviations (SD) below the mean to avoid floor effects and ceiling effects. Math tests having this quality can be used to identify students who have serious problems with math (i.e., who score at least 2 SD below the mean) and who are particularly strong in mathematics (i.e., score 2 SD above the mean). To examine this characteristic of validity, we converted the raw scores in the sample to standard scores having a mean of 100 and standard deviation of 15. Thus, the scores should range from 70 to 130 (+ or – 30, which is 2 SD) in order for floor ceiling

© 2008 The University of Texas System/Texas Education Agency

effects to be absent. We have set an arbitrary figure of 1.5 SD in order for the scores to "approximate" absence by approaching 2 SD.

If we were to rigidly follow Bracken and Rathvon, for kindergarten, only 27 of the 36 scores would meet the criterion for avoiding floor (12 of 18, 67%) and ceiling (15 of 18, 83%) effects. The biggest culprits, not surprisingly, occur in the fall, when testing kindergarteners is somewhat questionable because their scores are likely to be very low. But closer examination of the data provides a less gloomy conclusion. First, the Total Score of the TEMI-PM, which is the score used for identification purposes, approximates 2 SD (75 instead of 70) in the fall, and the TEMI-O score does not have floor effects. Second, 34 of 36 scores (94%) have scores that are more than 1.5 SD from the mean. For grade 1, 21 of 24 scores (88%) would meet the 2 SD criterion required for avoiding floor effects (all 24—100%—meet the 1.5 SD criterion); 20 of 24 (83%) scores would satisfy the 2 SD ceiling effects criterion; all would satisfy the 1.5 SD criterion. For grade 2, all scores meet the 2 SD criterion required for avoiding floor effects; 22 of 24 (92%) scores satisfy the 2 SD ceiling effects criterion. All but one (96%) satisfy the 1.5 SD ceiling criterion.

Table 137. Floor and Ceiling Effects of Kindergarten TEMI-PM and TEMI-O Measures

Test	Fall Floor	Fall Ceiling	Winter Floor	Winter Ceiling	Spring Floor	Spring Ceiling
MC	82	153	70	135	62	125
NID	73	145	52	160	47	149
NS	85	178	74	156	66	143
QR	71	161	59	157	52	152
Total PM	75	163	60	158	53	146
MPS O	54	130	36	127	23	123

Table 138. Floor and Ceiling Effects of Grade 1 TEMI-PM and TEMI-O Measures

Test	Fall Floor	Fall Ceiling	Winter Floor	Winter Ceiling	Spring Floor	Spring Ceiling
MC	67	149	52	144	48	134
NS	72	169	62	149	59	144
PV	72	198	62	151	52	158
ASC	81	172	70	156	63	142
Total PM	67	172	54	153	52	150
MPS	38	142	31	129	34	127
C	70	153	56	132	54	127
Total O	49	152	41	131	38	126

© 2008 The University of Texas System/Texas Education Agency

Table 139. Floor and Ceiling Effects of Grade 2 TEMI-PM and TEMI-O Measures

Test	Fall Floor	Fall Ceiling	Winter Floor	Winter Ceiling	Spring Floor	Spring Ceiling
MC	52	147	45	134	42	125
NS	56	204	52	163	58	151
PV	69	163	65	160	61	146
ASC	61	142	51	128	46	118
Total PM	56	169	56	151	52	139
MPS	41	142	27	140	48	133
C	53	150	40	143	46	132
Total O	42	150	32	145	41	134

Interrelationship of Mathematics Skills

Although math tests like the TEMI-PM and TEMI-O assess different components of math abilities, all of the abilities are numeric and should relate to one another. Thus, all of the subtests should intercorrelate to a moderate to high extent. Tables below demonstrate that the subtest scores of the TEMI-PM and TEMI-O do indeed intercorrelate to an expected level, thus providing additional support for the construct-identification validity of the tests' scores.

Table 140. Interrelationship Among TEMI-PM and TEMI-O Scores for Kindergarten in Fall

Score	TEMI-PM MC	TEMI-PM NID	TEMI-PM NID	TEMI-PM QR	TEMI-PM Total	TEMI-O MPS
TEMI-PM MC		.65	.69	.61	.84	.60
TEMI-PM NID			.67	.68	.79	.63
TEMI-PM NS				.63	.80	.60
TEMI-PM QR					.80	.57
TEMI-PM Total						.69

Table 141. Interrelationship Among TEMI-PM and TEMI-O Scores for Kindergarten in Winter

Score	TEMI-PM MC	TEMI-PM NID	TEMI-PM NS	TEMI-PM QR	TEMI-PM Total	TEMI-O MPS
TEMI-PM MC		.59	.61	.55	.81	.52
TEMI-PM NID			.61	.62	.74	.48
TEMI-PM NS				.61	.77	.56
TEMI-PM QR					.76	.37
TEMI-PM Total						.57

© 2008 The University of Texas System/Texas Education Agency

Table 142. Interrelationship Among TEMI-PM and TEMI-O Scores for Kindergarten in Spring

Score	TEMI-PM MC	TEMI-PM NID	TEMI-PM NS	TEMI-PM QR	TEMI-PM Total	TEMI-O MPS
TEMI-PM MC		.57	.60	.58	.81	.47
TEMI-PM NID			.65	.64	.74	.44
TEMI-PM NS				.70	.80	.44
TEMI-PM QR					.80	.33
TEMI-PM Total						.48

Table 143. Interrelationship Among TEMI-PM and TEMI-O Scores for Grade 1 in Fall

TEMI-PM	TEMI-PM MC	TEMI-PM NS	TEMI-PM PV	TEMI-PM ASC	TEMI-PM Total	TEMI-O MPS	TEMI-O C	TEMI-O Total
TEMI-PM MC		.67	.59	.59	.84	.49	.51	.57
TEMI-PM NS			.68	.64	.82	.49	.54	.59
TEMI-PM PV				.70	.77	.44	.52	.55
TEMI-PM ASC					.76	.45	.64	.62
TEMI-PM Total						.55	.63	.68
TEMI-O MPS							.52	.78
TEMI-O C								.80

© 2008 The University of Texas System/Texas Education Agency

Table 144. Interrelationship Among TEMI-PM and TEMI-O Scores for Grade 1 in the Winter

TEMI-PM	TEMI-PM MC	TEMI-PM NS	TEMI-PM PV	TEMI-PM ASC	TEMI-PM Total	TEMI-O MPS	TEMI-O C	TEMI-O Total
TEMI-PM MC		.70	.63	.63	.84	.49	.49	.55
TEMI-PM NS			.69	.67	.84	.51	.52	.58
TEMI-PM PV				.68	.80	.50	.52	.58
TEMI-PM ASC					.80	.51	.63	.65
TEMI-PM Total						.58	.62	.68
TEMI-O MPS							.56	.78
TEMI-O C								.80

Table 145. Interrelationship Among TEMI-PM and TEMI-O Scores for Grade 1 in the Spring

TEMI-PM	TEMI-PM MC	TEMI-PM NS	TEMI-PM PV	TEMI-PM ASC	TEMI-PM Total	TEMI-O MPS	TEMI-O C	TEMI-O Total
TEMI-PM MC		.70	.62	.61	.83	.43	.47	.50
TEMI-PM NS			.67	.65	.84	.43	.48	.51
TEMI-PM PV				.66	.79	.42	.49	.50
TEMI-PM ASC					.79	.46	.60	.59
TEMI-PM Total						.50	.59	.61
TEMI-O MPS							.61	.82
TEMI-O C								.80

© 2008 The University of Texas System/Texas Education Agency

Table 146. Interrelationship Among TEMI-PM and TEMI-O Scores for Grade 2 in the Fall

TEMI-PM	TEMI-PM MC	TEMI-PM NS	TEMI-PM PV	TEMI-PM ASC	TEMI-PM Total	TEMI-O MPS	TEMI-O C	TEMI-O Total
TEMI-PM MC		.51	.52	.50	.80	.42	.38	.47
TEMI-PM NS			.56	.65	.79	.39	.44	.50
TEMI-PM PV				.51	.71	.51	.43	.55
TEMI-PM ASC					.75	.37	.48	.51
TEMI-PM Total						.51	.52	.61
TEMI-O MPS							.38	.61
TEMI-O C								.77
TEMI-O Total								

Table 147. Interrelationship Among TEMI-PM and TEMI-O Scores for Grade 2 in the Winter

TEMI-PM	TEMI-PM MC	TEMI-PM NS	TEMI-PM PV	TEMI-PM ASC	TEMI-PM Total	TEMI-O MPS	TEMI-O C	TEMI-O Total
TEMI-PM MC		.58	.47	.55	.78	.45	.39	.47
TEMI-PM NS			.59	.67	.79	.43	.46	.51
TEMI-PM PV				.52	.67	.42	.42	.48
TEMI-PM ASC					.78	.47	.58	.61
TEMI-PM Total						.54	.56	.63
TEMI-O MPS							.50	.69
TEMI-O C								.83

© 2008 The University of Texas System/Texas Education Agency

Table 148. Interrelationship Among TEMI-PM and TEMI-O Scores for Grade 2 in the Spring

TEMI-PM	TEMI-PM MC	TEMI-PM NS	TEMI-PM PV	TEMI-PM ASC	TEMI-PM Total	TEMI-O MPS	TEMI-O C	TEMI-O Total
TEMI-PM MC		.61	.55	.55	.80	.43	.37	.45
TEMI-PM NS			.60	.60	.79	.40	.41	.46
TEMI-PM PV				.52	.71	.40	.37	.43
TEMI-PM ASC					.75	.42	.52	.54
TEMI-PM Total						.50	.50	.57
TEMI-O MPS							.53	.71
TEMI-O C								.85

To further examine the relationship among TEMI-PM and TEMI-O scores, an exploratory factor analysis with varimax rotation was conducted on the Kindergarten TEMI-PM (Magnitude Comparisons, Number Sequence, Number Identification, and Quantitative Reasoning) and TEMI-O (Mathematics Problem Solving) subtest scores to assess dimensionality. Based on the scree test and the "eigenvalue greater than one" rule, all five subtests were shown to load on one factor, which we refer to as "Mathematics Ability." Exploratory factor analyses were also conducted on the first- and second-grade TEMI-PM (Magnitude Comparisons, Number Sequence, Place Value, and Addition/Subtraction) and TEMI-O (Mathematics Problem Solving and Computation) subtest scores. At both grade levels, all six subtests loaded on a single factor (i.e., "Mathematics Ability"). The subtests that had the lowest factor loadings in the kindergarten, first-, and second-grade analyses were those contained within the TEMI-O. This is to be expected, given that the TEMI-O subtests are untimed, whereas the TEMI-PM includes only timed tests and therefore there is shared variance within each type of test.

Group Differences

Tests that are used to identify students with a particular condition (in this case, math difficulties) should produce test scores that are lower for students with known problems. To test this hypothesis, we looked at TEMI-PM and TEMI-O scores for students who scored below average on the SAT-10 (standard scores less than 90) and those who scored average or above (90 and higher on the SAT-10). The mean TEMI-PM and TEMI-O scores for the two groups are provided below. The substantial difference in scores shown in the tables, verified when *T*-tests showed differences significant beyond $p < .001$, provide additional support for the construct-identification validity of the TEMI-PM and TEMI-O scores.

© 2008 The University of Texas System/Texas Education Agency

Table 149. Means and Standard Deviations Showing Group Differences Between Lower and Higher Achievers, Fall Kindergarten

	SAT-10 Total < 90		SAT-10 Total > 89	
Score	Mean	Std. Dev.	Mean	Std. Dev.
TEMI-PM MC	6.47	6.977	15.83	12.847
TEMI-PM NID	1.76	2.475	6.64	6.394
TEMI-PM NS	4.24	3.754	8.15	4.362
TEMI-PM QR	13.24	9.235	22.74	10.334
TEMI-PM Total	25.71	18.145	53.35	29.042
TEMI-O MPS	12.52	6.013	15.83	12.847

Table 150. Means and Standard Deviations Showing Group Differences Between Lower and Higher Achievers, Winter Kindergarten

	SAT-10 Total < 90		SAT-10 Total > 89	
Score	Mean	Std. Dev.	Mean	Std. Dev.
TEMI-PM MC	20.55	12.036	19.66	6.247
TEMI-PM NID	7.94	5.931	32.89	13.119
TEMI-PM NS	9.63	3.638	13.73	6.505
TEMI-PM QR	22.63	8.683	13.15	3.046
TEMI-PM Total	60.75	24.481	29.66	8.894
TEMI-O MPS	20.00	4.703	89.43	25.633

Table 151. Means and Standard Deviations Showing Group Differences Between Lower and Higher Achievers, Spring Kindergarten

	SAT-10 Total < 90		SAT-10 Total > 89	
Score	Mean	Std. Dev.	Mean	Std. Dev.
TEMI-PM MC	31.04	14.788	42.66	12.992
TEMI-PM NID	12.27	6.449	18.74	7.018
TEMI-PM NS	12.72	3.781	15.46	3.343
TEMI-PM QR	28.52	9.108	33.47	8.026
TEMI-PM Total	84.55	26.242	110.34	25.572
TEMI-O MPS	21.91	4.747	27.29	3.642

Table 152. Means and Standard Deviations Showing Group Differences Between Lower and Higher Achievers, Fall Grade 1

	SAT-10 Total < 90		SAT-10 Total > 89	
Score	Mean	Std. Dev.	Mean	Std. Dev.
TEMI-PM MC	21.28	10.239	30.68	10.725
TEMI-PM NS	8.14	5.026	13.58	6.468
TEMI-PM PV	5.30	3.251	7.10	3.416
TEMI-PM ASC	3.16	2.357	6.37	4.269
TEMI-PM Total	37.88	17.097	57.72	20.772
TEMI-O MPS	17.54	4.290	22.01	4.817
TEMI-O C	7.20	3.616	11.80	4.918
TEMI-O Total	24.74	6.326	33.81	8.304

© 2008 The University of Texas System/Texas Education Agency

Table 153. Means and Standard Deviations Showing Group Differences Between Lower and Higher Achievers, Winter Grade 1

	SAT-10 Total < 90		SAT-10 Total > 89	
Score	Mean	Std. Dev.	Mean	Std. Dev.
TEMI-PM MC	29.95	8.791	37.35	9.098
TEMI-PM NS	13.41	5.978	19.34	6.419
TEMI-PM PV	9.62	4.549	13.38	4.524
TEMI-PM ASC	8.80	4.593	14.60	5.722
TEMI-PM Total	61.79	18.575	84.66	21.347
TEMI-O MPS	21.55	5.044	27.21	4.527
TEMI-O C	13.08	5.245	18.90	4.843
TEMI-O Total	34.62	8.858	46.11	7.821

Table 154. Means and Standard Deviations Showing Group Differences Between Lower and Higher Achievers, Spring Grade 1

	SAT-10 Total < 90		SAT-10 Total > 89	
Score	Mean	Std. Dev.	Mean	Std. Dev.
TEMI-PM MC	34.14	9.325	41.43	9.610
TEMI-PM NS	15.90	6.563	21.91	6.634
TEMI-PM PV	15.29	4.656	19.56	4.081
TEMI-PM ASC	13.80	5.000	20.34	6.877
TEMI-PM Total	79.12	20.027	103.09	21.986
TEMI-O MPS	23.70	5.354	30.80	3.470
TEMI-O C	15.81	4.308	22.59	3.667
TEMI-O Total	39.51	7.896	53.38	5.917

Table 155. Means and Standard Deviations Showing Group Differences Between Lower and Higher Achievers, Fall Grade 2

	SAT-10 Total < 90		SAT-10 Total > 89	
Score	Mean	Std. Dev.	Mean	Std. Dev.
TEMI-PM MC	29.52	9.924	38.90	9.431
TEMI-PM NS	8.57	3.451	12.23	4.315
TEMI-PM PV	4.86	2.777	8.92	3.385
TEMI-PM ASC	11.95	5.724	19.90	7.492
TEMI-PM Total	54.90	16.376	79.94	20.103
TEMI-O MPS	17.52	4.636	23.80	4.612
TEMI-O C	16.92	6.027	27.19	7.811
TEMI-O Total	34.43	8.014	50.98	10.644

© 2008 The University of Texas System/Texas Education Agency

Table 156. Means and Standard Deviations Showing Group Differences Between Lower and Higher Achievers, Winter Grade 2

	SAT-10 Total < 90		SAT-10 Total > 89	
Score	Mean	Std. Dev.	Mean	Std. Dev.
TEMI-PM MC	34.8810	8.30649	43.9086	8.87478
TEMI-PM NS	13.3855	4.70296	17.4195	4.81350
TEMI-PM PV	8.8095	4.01646	11.3600	3.43438
TEMI-PM ASC	17.2738	6.44598	26.1086	7.32298
TEMI-PM Total	74.1446	16.96778	98.7931	19.57564
TEMI-O MPS	21.0119	4.08492	27.3257	3.77258
TEMI-O C	21.7857	6.24610	32.0914	6.09529
TEMI-O Total	42.7976	8.38929	59.4171	8.61653

Table 157. Means and Standard Deviations Showing Group Differences Between Lower and Higher Achievers, Spring Grade 2

	SAT-10 Total < 90		SAT-10 Total > 89	
Score	Mean	Std. Dev.	Mean	Std. Dev.
TEMI-PM MC	41.2619	10.08413	51.3864	9.02243
TEMI-PM NS	15.5595	5.15156	20.3239	5.30043
TEMI-PM PV	10.7619	5.43886	14.5795	4.36145
TEMI-PM ASC	24.4643	7.04965	31.5227	7.64868
TEMI-PM Total	92.0476	20.68752	117.8125	21.04157
TEMI-O MPS	21.4643	4.67613	29.2171	3.30947
TEMI-O C	28.0833	6.08020	37.8114	6.17130
TEMI-O Total	49.5476	9.13804	67.0286	7.72251

Table 158. Tests for Equality of Means Between Lower and Higher Achievers, Spring Kindergarten

	Levene's Test for Equality of Variances			t-test for Equality of Means		
TEMI Score		F	Sig.	t	df	Sig. (2-tailed)
MC Fall	Equal variances not assumed	32.692	.000	-7.444	207.838	.000
NID Fall	Equal variances assumed	.907	.342	-6.505	260	.000
NS Fall	Equal variances not assumed	67.899	.000	-8.890	255.427	.000
QR Fall	Equal variances assumed	.153	.696	-6.627	260	.000
Total PM Fall	Equal variances not assumed	16.123	.000	-9.067	180.685	.000
MPS O Fall	Equal variances assumed	.003	.954	-8.109	259	.000
MC Winter	Equal variances assumed	.267	.606	-6.786	262	.000
NID Winter	Equal variances assumed	1.030	.311	-7.765	262	.000
NS Winter	Equal variances assumed	.674	.412	-6.433	262	.000
QR Winter	Equal variances assumed	.610	.435	-5.629	262	.000
Total PM Winter	Equal variances assumed	.022	.884	-8.002	262	.000
MPS O Winter	Equal variances assumed	.344	.558	-9.916	261	.000
MC Spring	Equal variances assumed	1.811	.180	-6.104	263	.000
NID Spring	Equal variances assumed	.661	.417	-5.623	263	.000
NS Spring	Equal variances assumed	.175	.676	-6.658	263	.000
QR Spring	Equal variances assumed	3.649	.057	-4.216	263	.000
Total PM Spring	Equal variances assumed	.554	.457	-7.089	263	.000
MPS O Spring	Equal variances not assumed	6.983	.009	-8.467	93.656	.000

© 2008 The University of Texas System/Texas Education Agency

Table 159. Tests for Equality of Means Between Lower and Higher Achievers, Spring Grade 1

	Levene's Test for Equality of Variances			t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
MC Fall	Equal variances assumed	.188	.665	-6.783	295	.000
NS Fall	Equal variances assumed	4.575	.033	-6.800	295	.000
PV Fall	Equal variances assumed	3.179	.076	-4.073	295	.000
ASC Fall	Equal variances assumed	25.417	.000	-6.365	295	.000
Total PM Fall	Equal variances assumed	2.486	.116	-7.641	295	.000
MPS Fall	Equal variances assumed	.137	.711	-7.302	295	.000
C Fall	Equal variances assumed	11.985	.001	-7.631	295	.000
Total O Fall	Equal variances assumed	6.878	.009	-8.862	295	.000
MC Winter	Equal variances assumed	.588	.444	-6.272	294	.000
NS Winter	Equal variances not assumed	.018	.892	-7.421	150.766	.000
PV Winter	Equal variances not assumed	.006	.940	-6.308	140.553	.000
ASC Winter	Equal variances assumed	2.090	.149	-8.139	294	.000
Total PM Winter	Equal variances assumed	1.672	.197	-8.468	294	.000
MPS Winter	Equal variances assumed	1.504	.221	-9.237	292	.000
C Winter	Equal variances assumed	.503	.479	-8.974	292	.000
Total O Winter	Equal variances assumed	.804	.371	-10.798	292	.000
MC Spring	Equal variances assumed	.696	.405	-5.844	294	.000
NS Spring	Equal variances assumed	.152	.697	-6.939	294	.000
PV Spring	Equal variances assumed	4.346	.038	-7.685	294	.000
ASC Spring	Equal variances assumed	6.293	.013	-7.777	293	.000
Total PM Spring	Equal variances assumed	.521	.471	-8.522	293	.000
MPS Spring	Equal variances assumed	24.331	.000	-13.360	295	.000
C Spring	Equal variances assumed	2.016	.157	-13.452	295	.000
Total O Spring	Equal variances assumed	13.082	.000	-16.299	295	.000

© 2008 The University of Texas System/Texas Education Agency

Table 160. Tests for Equality of Means Between Lower and Higher Achievers, Spring Grade 2

	Levene's Test for Equality of Variances			t-test for Equality of Means		
		F	Sig.	T	df	Sig. (2-tailed)
MC Fall	Equal variances assumed	.228	.633	-7.369	258	.000
NS Fall	Equal variances assumed	3.492	.063	-6.794	258	.000
PV Fall	Equal variances assumed	.900	.344	-9.569	258	.000
ASC Fall	Equal variances not assumed	7.380	.007	-9.436	208.184	.000
Total PM Fall	Equal variances assumed	1.894	.170	-9.946	258	.000
MPS Fall	Equal variances assumed	.684	.409	-10.204	257	.000
C Fall	Equal variances assumed	3.508	.062	-10.583	257	.000
Total O Fall	Equal variances not assumed	4.809	.029	-13.900	207.790	.000
MC Winter	Equal variances assumed	.627	.429	-7.822	257	.000
NS Winter	Equal variances assumed	.011	.915	-6.329	255	.000
PV Winter	Equal variances not assumed	4.475	.035	-5.007	143.075	.000
ASC Winter	Equal variances assumed	.520	.472	-9.439	257	.000
Total PM Winter	Equal variances assumed	2.684	.103	-9.841	255	.000
MPS Winter	Equal variances assumed	.568	.452	-12.271	257	.000
C Winter	Equal variances assumed	.084	.772	-12.636	257	.000
Total O Winter	Equal variances assumed	.027	.871	-14.655	257	.000
MC Spring	Equal variances assumed	.481	.488	-8.142	258	.000
NS Spring	Equal variances assumed	.013	.909	-6.839	258	.000
PV Spring	Equal variances assumed	3.387	.067	-6.080	258	.000
ASC Spring	Equal variances assumed	1.150	.284	-7.134	258	.000
Total PM Spring	Equal variances assumed	.105	.747	-9.283	258	.000
MPS Spring	Equal variances not assumed	14.052	.000	-13.644	124.282	.000
C Spring	Equal variances assumed	.119	.731	-11.932	257	.000
Total O Spring	Equal variances assumed	2.934	.088	-16.048	257	.000

Relationship to Written Language

Although there are students who exhibit typical strengths and weaknesses across academic domains (that is, some students do better at reading than math, or do better in math than writing or social studies), these differences are relative. Most students who do well in some academic areas tend to do fairly well in other academic subjects. Thus, test scores in math should relate to test scores in other academic subjects.

We correlated the results of the TEMI-PM and TEMI-O with that of two measures of written language. First, we correlated our measures' scores with those of the ITBS reading and spelling scores. Second, we correlated the TEMI-PM and TEMI-O test scores with those of the Test of Silent Reading Efficiency (TOSRE; Wagner, Torgesen, Rashotte, & Pearson, in press).

In the second study, we administered the TOSRE to first- and second-graders in the spring of 2008. Correlation coefficients examining the construct validity of the TEMI-PM and TEMI-O test scores are reported below.

© 2008 The University of Texas System/Texas Education Agency

Table 161. Correlation Coefficients Depicting the Relationship Between TEMI-PM and TEMI-O Scores with ITBS Reading Scores—Kindergarten

ITBS Score	TEMI-PM MC	TEMI-PM NI	TEMI-PM NS	TEMI-PM QR	TEMI-PM Total	TEMI-O MPS
Vocabulary	.416	.342	.421	.308	.428	.445
Word Analysis	.585	.559	.610	.499	.643	.488

Table 162. Correlation Coefficients Depicting the Relationship Between TEMI-PM and TEMI-O Scores with ITBS Reading Scores—Grade 1

ITBS Score	TEMI-PM MC	TEMI-PM NS	TEMI-PM ASC	TEMI-PM PV	TEMI-PM Total	TEMI-O MPS	TEMI-O C	TEMI-O Total
Vocabulary	.334	.277	.226	.282	.341	.428	.397	.457
Words	.497	.586	.492	.497	.612	.519	.583	.633
Comprehension	.387	.508	.462	.462	.525	.468	.518	.576
Reading Total	.477	.591	.505	.504	.609	.543	.588	.651
Word Analysis	.482	.612	.420	.548	.611	.593	.556	.654
Language	.311	.438	.307	.400	.421	.539	.483	.575

Table 163. Correlation Coefficients Depicting the Relationship Between TEMI-PM and TEMI-O Scores with ITBS Reading Scores—Grade 2

ITBS Score	TEMI-PM MC	TEMI-PM NS	TEMI-PM ASC	TEMI-PM PV	TEMI-PM Total	TEMI-O MPS	TEMI-O C	TEMI-O Total
Vocabulary	.182	.175	.290	.300	.271	.416	.267	.380
Comprehension	.338	.302	.336	.338	.385	.489	.325	.455
Reading Total	.261	.241	.329	.331	.337	.468	.302	.428
Word Analysis	.280	.351	.327	.380	.377	.441	.321	.433
Spelling	.229	.335	.323	.373	.350	.339	.221	.312
Language Total	.221	.271	.309	.408	.335	.497	.350	.480

The coefficients reported above provide support for the construct validity of the test scores reported for the TEMI-PM and TEMI-O.

Relationship to Total Achievement

In the previous section, we noted that math scores should relate to other scores of academic achievement (i.e., reading and writing). Here, we extend that hypothesis to overall academic skills. That is, as measures of achievement, the TEMI-PM and TEMI-O should relate to total scores of achievement tests. To test this hypothesis, we correlated our test results to those of the Total (Core) Score of the ITBS (see tables below).

Table 164. Relationship Between TEMI-PM and TEMI-O Scores with ITBS Core (Total) Scores—Kindergarten

ITBS Score	TEMI-PM MC	TEMI-PM NI	TEMI-PM NS	TEMI-PM QR	TEMI-PM Total	TEMI-O MPS
Core total	.60	.57	.51	.48	.63	.57

© 2008 The University of Texas System/Texas Education Agency

Table 165. Relationship Between TEMI-PM and TEMI-O Scores with ITBS Core (Total) Scores—Grade 1

ITBS Score	TEMI-PM MC	TEMI-PM NS	TEMI-PM ASC	TEMI-PM PV	TEMI-PM Total	TEMI-O MPS	TEMI-O C	TEMI-O Total
Core total	.44	.49	.41	.49	.54	.61	.58	.66

Table 166. Relationship Between TEMI-PM and TEMI-O Scores with ITBS Core (Total) Scores—Grade 2

ITBS	TEMI-PM MC	TEMI-PM NS	TEMI-PM ASC	TEMI-PM PV	TEMI-PM Total	TEMI-O MPS	TEMI-O C	TEMI-O Total
Core Total	.35	.32	.42	.48	.45	.60	.41	.57

Item Validity

Guilford and Fruchter (1978) noted that correlations between items and the total score of the test they contribute to provide evidence of item validity. We reported discriminating powers, an index of item validity, in an earlier section of this manual. The magnitude of these coefficients provides support for the validity of the TEMI-PM and TEMI-O items and, consequently, for the construct-identification validity of the TEMI-PM and TEMI-O test scores.

© 2008 The University of Texas System/Texas Education Agency

References

Aiken, L. R. (2000). *Psychological testing and assessment* (8th ed.). Needham Heights, MA: Allyn & Bacon.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). New York: Macmillan.

Benoit, L., Lehalle, H., & Jouen, F. (2004). Do young children acquire number words through subitizing or counting? *Cognitive Development, 19*(3), 291–307.

Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities, 38*(4), 333–339.

Bracken, B. A. (1987). Limitations of preschool instruments and standards for minimal technical adequacy. *Journal of Psychoeducational Assessment, 4*, 313–326.

Camilli, G., & Shepard, L. (1994). *Methods for identifying biased test items.* Thousand Oaks, CA: Sage.

Chard, D. J., Clarke, B., Baker, S., Otterstedt, J., Braun, D., & Katz, R. (2005). Using measures of number sense to screen for difficulties in mathematics: Preliminary findings. *Assessment for Effective Intervention, 30*(2), 3–14.

Clark, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33*(2), 234–238.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297–334.

Dehaene, S. (1997). *The number sense*. New York: Oxford University.

Ebel, R. L. (1972). Essentials of educational measurement (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97,* 493–513.

Fuchs, L.S., Fuchs, D., Hamlet, C. L., Powell, S. R., Capizzi, A. M., & Seethaler, P. M. (2006). The effects of computer-assisted instruction on number combination skill in at-risk first graders. *Journal of Learning Disabilities, 39*(5), 467–475.

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Lipsey, M. W. (2002). Reading differences between low-achieving students with and without learning disabilities: A meta-analysis. In R. Gersten, E. Schiller, & S. Vaughn (Eds.), *Contemporary special education research* (pp. 81–104). Mahwah, NJ: Lawrence Erlbaum Associates.

Garrett, H. (1965). *Testing for teachers.* New York: American Book.

Guilford, J. P., & Fruchter, B. (1978). *Fundamental statistics in psychology and education.* New York: McGraw-Hill.

Hammill, D. D., & Bryant, B. R. (1998). *Learning disability diagnostic inventory*. Austin, TX: PRO-ED.

Hammill, D. D., & Bryant, B. R. (2005). *Detroit tests of learning aptitude-primary: Third edition*. Austin, TX: PRO-ED.

Harcourt Assessment. (2003). *Stanford achievement test: Tenth edition.* San Antonio, TX: Author.

Ho Yu, C. (Undated). *An introduction to computing and interpreting Cronbach Coefficient Alpha in SAS*. Retrieved September 12, 2008, from http://psyphz.psych.wisc.edu/~shackman/cronbach.doc.

© 2008 The University of Texas System/Texas Education Agency

Jodoin, M. G., & Gierl, M. J. (2000, April). *Evaluating Type I error rates using an effect size measure with the logistic regression procedure for DIF detection.* Paper presented at the meeting of the American Educational Research Association, New Orleans.

Jodoin, M. G., & Gierl, M. J. (2001). Evaluating Type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education, 14*, 329–349.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology, 62,* 498–525.

Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching* (8th ed.). Upper Saddle River, NJ: Prentice-Hall.

Miller, T. R., & Spray, J. A. (1993). Logistic discriminant function analysis for DIF identification of polytomously scored items. *Journal of Educational Measurement, 30*, 107–122.

Newcomer, P. (2008). *Diagnostic Achievement Battery – 4th edition*. Austin, TX: PRO-ED.

Oosterhof, A. C. (1976). Similarity of various item discrimination indices. *Journal of Educational Measurement, 13*, 145–150.

Pyrczak, F. (1973). Validity of the discrimination index as a measure of item validity. *Journal of Educational Measurement, 10*, 227–231.

Rathvon, N. (2004). *Early reading assessment: A practitioner's handbook*. New York: Guilford.

Rozeboom, W. W. (1966). Foundation of the theory of prediction. Homewood, IL: Dorsey.

Salvia, J., & Ysseldyke, J. (2001). *Assessment in special and remedial education* (8th ed.). Boston: Houghton Mifflin.

Smith, D. D., & Lovitt, T. C. (1982). Computational arithmetic program. Austin, TX: PRO-ED.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 26*, 55–66.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (in press). *Test of silent reading efficiency and comprehension*. Austin, TX: PRO-ED.

Zumbo, B. D. (1999). A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

© 2008 The University of Texas System/Texas Education Agency