

# 5

---

## Psychometric Properties

This chapter presents data to support the reliability and validity of the TOD. Chapter 4 described the preliminary development of the TOD tests and Rating Scales and related pilot data, along with the demographic composition of the standardization and clinical samples. Analyses presented in this chapter were derived from the standardization and clinical samples described in Chapter 4.

---

### Reliability

*Reliability* refers to the accuracy or precision of test scores. Reliability coefficients capture the extent to which the results are dependable and relatively free from error. The standard error of measurement, or *SEM*, is derived from statistical estimates of reliability and is frequently used to indicate the precision characterizing an individual score. The smaller the *SEM*, the higher the reliability. This section presents evidence that the TOD test, index, and composite scores are sufficiently reliable and precise for measuring an individual's skills.

This first section of the chapter offers a review of several reliability concepts and a description of different types of reliability analyses performed for the TOD tests, indexes, composites, and Rating Scales.

#### **Internal Consistency**

*Internal consistency* refers to the extent to which all items in a test or scale consistently measure the same

ability or trait. Internal consistency is estimated as a reliability coefficient, which ranges from 0 to 1. The methodology for estimating an internal consistency coefficient depends upon the type of test. Because of the different types of tests within the TOD, various methodologies were used to estimate internal consistency estimates, as described in the following paragraphs.

#### ***Split-Half Reliability***

Internal consistency reliability for most of the TOD tests was calculated using the *split-half method* (Cronbach, 1970). This procedure involves splitting test items into halves based on their difficulty. Raw scores from the two halves were correlated using the Pearson product-moment correlation coefficient, and then adjusted using the Spearman-Brown formula (Anastasi & Urbina, 1997).

### **Rasch-Based Reliability**

*Rasch-based reliability* was used for two different types of tests: tests using item sets and speeded (timed) tests. The item set format of Picture Vocabulary (1S) and Letter and Word Choice (2S) involved Rasch methodology to derive person ability scores used to calculate internal consistency. Internal consistency reliability for timed tests should not be calculated using traditional methods because items are sensitive to *both* speed and accuracy. Therefore, the Rasch model was applied to calculate reliability for all speeded tests.

The Rasch analyses yield an estimated ability score on a logit scale and standard error (*SE*) for each person in the standardization sample. Each person's *SE* is then squared to produce an error variance estimate, and the mean error variance ( $SE^2$ ) and the variance of the Rasch ability estimate ( $SD^2$ ) are computed for all persons. Reliability is then estimated as  $r = 1 - (SE^2 / SD^2)$ .

### **Coefficient Alpha**

*Coefficient alpha* (Cronbach, 1988) is the most frequently used methodology for estimating reliability for rating scales (and other tests without a developmental gradient) and was applied for the TOD Rating Scales. It represents a more conservative estimate of reliability.

### **Reliability of Linear Combinations**

All indexes and composites were created by summing tests, and therefore internal consistency reliability for these scores was estimated using the formula for reliability of linear combinations (Nunnally & Bernstein, 1994).

Tables 5.1 and 5.2 show the internal consistency coefficients for the TOD-S tests and Dyslexia Risk Index (DRI) in the TOD-S child and adult standardization samples, respectively. Internal consistency estimates in Table 5.1 are provided by grade because the selection of TOD-S form in children up to age 18 is determined by grade and not age, whereas those in Table 5.2 are displayed by age range because all

adults take the same form. Results illustrate that all reliability coefficients are  $\geq .70$ , with most  $\geq .80$ , indicating acceptable levels of internal consistency that support the use of these scores in clinical applications.

Table 5.3 shows the internal consistency estimates for TOD-C tests in a combined child and adult standardization sample. Internal consistency cannot be calculated for the Oral Reading Efficiency (12C) test because the test functions psychometrically as a single "item." Most reliability coefficients are  $\geq .90$ , and almost all are  $> .80$ . Less than .02% of analyses yielded reliabilities lower than .70.

Tables 5.4 and 5.5 display internal consistency estimates for the Dyslexia Diagnostic Index (DDI), Reading and Spelling Index (RSI), and Linguistic Processing Index (LPI), as well as the composites that can be calculated using the TOD-S and TOD-C tests, for the child and adult samples, respectively. Composite score reliabilities are displayed by grade for the child sample because TOD-S tests, which are administered by grade, are included in most of the composites. Reliabilities for adults are displayed by age range. All composite reliabilities are  $> .80$ , and most are  $> .90$ , indicating excellent reliability.

Table 5.6 displays the internal consistency estimates for the three TOD-C Rating Scales (Self, Parent/Caregiver, and Teacher). All are  $> .90$ , indicating excellent reliability.

Tables 5.7 and 5.8 display the internal consistency estimates for the TOD-E tests, indexes, and composites, and Table 5.9 displays the internal consistency estimates for the two TOD-E Rating Scales (Parent/Caregiver and Teacher). Almost all internal consistency estimates are  $> .90$ , and all are  $> .80$ . These results suggest excellent reliability for the TOD-E and support the use of these scores in clinical applications.

These same reliability analyses were performed on the TOD clinical samples, and all internal consistency reliabilities for the clinical samples were consistent with those from the standardization samples.

**Table 5.1.** TOD-S Internal Consistency Estimates and SEMs for Tests and Indexes by Grade: Child

Grade	n	Tests								Indexes			
		PV-S		LWC-S		WRF-S		QRF-S		DRI (WRF)		DRI (QRF)	
		r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM
K	121	.80	6.7	.85	5.8	.93	4.0	—	—	.93	4.0	—	—
1	199	.82	6.4	.88	5.2	.95	3.4	—	—	.88	5.2	—	—
2	221	.74	7.6	.85	5.8	—	—	.97	2.6	—	—	.85	5.8
3	170	.72	7.9	.81	6.5	—	—	.97	2.6	—	—	.93	4.0
4	146	.78	7.0	.83	6.2	—	—	.97	2.6	—	—	.83	6.2
5	140	.83	6.2	.81	6.5	—	—	.98	2.1	—	—	.81	6.5
6	145	.72	7.9	.84	6.0	—	—	.98	2.1	—	—	.95	3.4
7	128	.74	7.6	.81	6.5	—	—	.98	2.1	—	—	.94	3.7
8	103	.72	7.9	.78	7.0	—	—	.98	2.1	—	—	.92	4.2
9	101	.70	8.2	.78	7.0	—	—	.98	2.1	—	—	.92	4.2
10	88	.81	6.5	.83	6.2	—	—	.99	1.5	—	—	.94	3.7
11	87	.80	6.7	.85	5.8	—	—	.99	1.5	—	—	.95	3.4
12	74	.83	6.2	.87	5.4	—	—	.98	2.1	—	—	.94	3.7

Note.  $N = 1,723$ . TOD-S Child internal consistency estimates were calculated by grade because selection of TOD-S form is based on grade (i.e., Grades K-1, 2-5, and 6-Adult). Internal consistency estimates for all tests were calculated using Rasch-based reliability. Reliability estimates for DRI (WRF) and DRI (QRF) were calculated using the reliability of linear combinations (Nunnally and Bernstein, 1994).  $SEM = SD \sqrt{1 - r}$ , where  $SEM$  is the standard error of measurement,  $SD$  is the standard deviation of the standard score unit (15), and  $r$  is the reliability coefficient. **PV-S** = Picture Vocabulary; **LWC-S** = Letter and Word Choice; **WRF-S** = Word Reading Fluency; **QRF-S** = Question Reading Fluency; **DRI (WRF)** = Dyslexia Risk Index (Word Reading Fluency); **DRI (QRF)** = Dyslexia Risk Index (Question Reading Fluency).

**Table 5.2.** TOD-S Internal Consistency Estimates and SEMs for Tests and Index by Age Range: Adult

Age (years)	n	Tests						Index	
		PV-S		LWC-S		QRF-S		DRI (QRF)	
		r	SEM	r	SEM	r	SEM	r	SEM
18-23	113	.77	7.2	.78	7.0	.99	1.5	.90	4.7
24-39	64	.72	7.9	.70	8.2	.99	1.5	.89	5.0
40-49	40	.74	7.6	.74	7.6	.99	1.5	.85	5.8
50-59	54	.74	7.6	.74	7.6	.99	1.5	.93	4.0
60-69	37	.86	5.6	.84	6.0	.99	1.5	.94	3.7
70-89	39	.88	5.2	.87	5.4	.99	1.5	.93	4.0

Note.  $N = 347$ . Internal consistency estimates for all tests were calculated using Rasch-based reliability. Reliability estimate for DRI (QRF) was calculated using the reliability of linear combinations (Nunnally and Bernstein, 1994).  $SEM = SD \sqrt{1 - r}$ , where  $SEM$  is the standard error of measurement,  $SD$  is the standard deviation of the standard score unit (15), and  $r$  is the reliability coefficient. **PV-S** = Picture Vocabulary; **LWC-S** = Letter and Word Choice; **QRF-S** = Question Reading Fluency; **DRI (QRF)** = Dyslexia Risk Index (Question Reading Fluency).

Table 5.3. TOD-C Internal Consistency Estimates and SEMs for Tests by Age Range: Child and Adult

Age (years)	n	PHM-C		IWS-C		RLN-C <sup>a</sup>		PWR-C		WPC-C <sup>a</sup>		WM-C		PAN-C		IWR-C		BLN-C		SEG-C	
		r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM
6	48	.94	3.7	.95	3.4	.97	2.6	.97	2.6	.97	2.6	.75	7.5	.91	4.5	.96	3.1	.65	8.9	.94	3.7
7	81	.96	3.0	.96	3.0	.99	1.5	.97	2.6	.99	1.5	.71	8.1	.85	5.8	.96	3.0	.81	6.5	.88	5.2
8	168	.92	4.2	.96	3.0	.98	2.1	.96	3.0	.99	1.5	.71	8.1	.91	4.5	.90	4.7	.82	6.4	.85	5.8
9	164	.91	4.5	.96	3.0	.98	2.1	.93	4.0	.99	1.5	.70	8.2	.88	5.2	.85	5.8	.87	5.4	.86	5.6
10	132	.89	5.0	.93	4.0	.99	1.5	.94	3.7	.98	2.1	.72	7.9	.82	6.4	.82	6.4	.87	5.4	.88	5.2
11	139	.88	5.2	.90	4.7	.98	2.1	.93	4.0	.99	1.5	.80	6.7	.85	5.8	.88	5.2	.90	4.7	.84	6.0
12	146	.88	5.2	.91	4.5	.99	1.5	.89	5.0	.99	1.5	.75	7.5	.91	4.5	.87	5.4	.86	5.6	.84	6.0
13	121	.82	6.4	.94	3.7	.94	3.7	.91	4.5	.99	1.5	.81	6.5	.85	5.8	.94	3.7	.92	4.2	.93	4.0
14	98	.85	5.8	.87	5.4	.99	1.5	.91	4.5	.99	1.5	.78	7.0	.68	8.5	.92	4.2	.90	4.7	.91	4.5
15	106	.74	7.6	.90	4.7	.99	1.5	.76	7.3	.99	1.5	.78	7.0	.88	5.2	.92	4.2	.86	5.6	.91	4.5
16	82	.74	7.6	.89	5.0	.99	1.5	.87	5.4	.98	2.1	.68	8.5	.88	5.2	.88	5.2	.92	4.2	.91	4.5
17	78	.81	6.5	.93	4.0	.99	1.5	.87	5.4	.96	3.0	.87	5.4	.84	6.0	.94	3.7	.94	3.7	.90	4.7
18	52	.87	5.4	.92	4.2	.99	1.5	.87	5.4	.99	1.5	.87	5.4	.78	7.0	.94	3.7	.87	5.4	.96	3.0
19-23	99	.83	6.2	.91	4.5	.81	6.5	.85	5.8	.98	2.1	.79	6.9	.84	6.0	.94	3.7	.94	3.7	.89	5.0
24-39	64	.77	7.2	.93	4.0	.99	1.5	.84	6.0	.99	1.5	.71	8.1	.82	6.4	.91	4.5	.90	4.7	.87	5.4
40-49	40	.87	5.4	.88	5.2	.99	1.5	.67	8.6	.98	2.1	.83	6.2	.78	7.0	.92	4.2	.82	6.4	.91	4.5
50-59	54	.76	7.3	.93	4.0	.99	1.5	.85	5.8	.98	2.1	.83	6.2	.87	5.4	.91	4.5	.93	4.0	.92	4.2
60-69	37	.93	4.0	.91	4.5	.99	1.5	.91	4.5	.99	1.5	.86	5.6	.85	5.8	.95	3.4	.95	3.4	.94	3.6
70-89	39	.94	3.7	.95	3.4	.99	1.5	.93	4.0	.99	1.5	.83	6.2	.94	3.7	.97	2.6	.97	2.6	.96	3.0

Note. N = 1,748. Internal consistency estimates for timed tests were calculated using Rasch-based reliability; all others were based on the split-half method.  $SEM = SD \sqrt{1 - r}$ , where SEM is the standard error of measurement, SD is the standard deviation of the standard score unit (15), and r is the reliability coefficient. PHM-C = Phonological Manipulation; IWS-C = Irregular Word Spelling; RLN-C = Rapid Letter Naming; PWR-C = Pseudoword Reading; WPC-C = Word Pattern Choice; WM-C = Word Memory; PAN-C = Picture Analogies; IWR-C = Irregular Word Reading; BLN-C = Blending; SEG-C = Segmenting; RWS-C = Regular Word Spelling; SREI-C = Silent Reading Efficiency Grades 1-5; SRE2-C = Silent Reading Efficiency Grade 6-Adult; RNL-C = Rapid Number and Letter Naming; LM-C = Letter Memory; RPW-C = Rapid Pseudoword Reading; RIW-C = Rapid Irregular Word Reading; SSL-C = Symbol to Sound Learning; LV-C = Listening Vocabulary; GAN-C = Geometric Analogies.  
<sup>a</sup>Timed test.

Table 5.3. TOD-C Internal Consistency Estimates and SEMs for Tests by Age Range: Child and Adult (continued)

Age (years)	n	RWS-C		SREI/2-C <sup>a</sup>		RNL-C <sup>a</sup>		LM-C		RPW-C <sup>a</sup>		RIW-C <sup>a</sup>		SSL-C		LV-C		GAN-C	
		r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM
6	48	.91	4.5	.98	2.1	.99	1.5	.71	8.1	.93	4.0	.96	3.0	.95	3.4	.84	6.0	.84	6.0
7	81	.94	3.7	.98	2.1	.99	1.5	.77	7.2	.96	3.0	.99	1.5	.93	4.0	.80	6.7	.93	4.0
8	168	.93	4.0	.98	2.1	.99	1.5	.61	9.4	.95	3.4	.98	2.1	.94	3.7	.82	6.4	.93	4.0
9	164	.96	3.0	.98	2.1	.96	3.0	.80	6.7	.94	3.7	.98	2.1	.95	3.4	.75	7.5	.93	4.0
10	132	.94	3.7	.95	3.4	.99	1.5	.75	7.5	.95	3.4	.98	2.1	.95	3.4	.82	6.4	.89	5.0
11	139	.94	3.7	.89	5.0	.99	1.5	.77	7.2	.96	3.0	.99	1.5	.96	3.0	.82	6.4	.90	4.7
12	146	.94	3.7	.96	3.0	.98	2.1	.80	6.7	.95	3.4	.97	2.6	.95	3.4	.83	6.2	.91	4.5
13	121	.92	4.2	.99	1.5	.97	2.6	.69	8.4	.94	3.7	.98	2.1	.96	3.0	.85	5.8	.89	5.0
14	98	.92	4.2	.98	2.1	.98	2.1	.73	7.8	.96	3.0	.97	2.6	.95	3.4	.88	5.2	.91	4.5
15	106	.87	5.4	.98	2.1	.99	1.5	.70	8.2	.97	2.6	.92	4.2	.95	3.4	.87	5.4	.88	5.2
16	82	.92	4.2	.99	1.5	.99	1.5	.77	7.2	.97	2.6	.95	3.4	.96	3.0	.91	4.5	.89	5.0
17	78	.92	4.2	.99	1.5	.99	1.5	.84	6.0	.96	3.0	.97	2.6	.98	2.1	.91	4.5	.88	5.2
18	52	.94	3.7	.95	3.4	.99	1.5	.81	6.5	.97	2.6	.96	3.0	.97	2.6	.85	5.8	.92	4.2
19-23	99	.90	4.7	.98	2.1	.99	1.5	.77	7.2	.96	3.0	.95	3.4	.94	3.7	.85	5.8	.88	5.2
24-39	64	.82	6.4	.97	2.6	.99	1.5	.84	6.0	.93	4.0	.94	3.7	.97	2.6	.86	5.6	.86	5.6
40-49	40	.88	5.2	.97	2.6	.99	1.5	.84	6.0	.94	3.7	.80	6.7	.96	3.0	.73	7.8	.88	5.2
50-59	54	.92	4.2	.97	2.6	.99	1.5	.84	6.0	.97	2.6	.95	3.4	.96	3.0	.87	5.4	.95	3.4
60-69	37	.91	4.5	.97	2.6	.99	1.5	.89	5.0	.94	3.7	.94	3.7	.99	1.5	.85	5.8	.91	4.5
70-89	39	.94	3.7	.98	2.1	.99	1.5	.58	9.7	.89	5.0	.94	3.7	.97	2.6	.94	3.7	.95	3.4

Note: N = 1,748. Internal consistency estimates for timed tests were calculated using Rasch-based reliability; all others were based on the split-half method.  $SEM = SD \sqrt{1 - r}$ , where SEM is the standard error of measurement, SD is the standard deviation of the standard score unit (15), and r is the reliability coefficient. **PHM-C** = Phonological Manipulation; **IWS-C** = Irregular Word Spelling; **RLN-C** = Rapid Letter Naming; **PWR-C** = Pseudoword Reading; **WPC-C** = Word Pattern Choice; **WM-C** = Word Memory; **PAN-C** = Picture Analogies; **IWR-C** = Irregular Word Reading; **BLN-C** = Blending; **SEG-C** = Segmenting; **RWS-C** = Regular Word Spelling; **SREI-C** = Silent Reading Efficiency Grades 1-5; **SRE2-C** = Silent Reading Efficiency Grade 6-Adult; **RNL-C** = Rapid Number and Letter Naming; **LM-C** = Letter Memory; **RPW-C** = Rapid Pseudoword Reading; **RIW-C** = Rapid Irregular Word Reading; **SSL-C** = Symbol to Sound Learning; **LV-C** = Listening Vocabulary; **GAN-C** = Geometric Analogies.

<sup>a</sup>Timed test.

Table 5.4. TOD-C Internal Consistency Estimates and SEMs for Indexes and Composites by Grade: Child

Grade	n	Indexes										Composites											
		DDI (WRF)		DDI (QRF)		LPI		RSI (WRF)		RSI (QRF)		SWA		PK		BRS		DE		SP		RF (WRF)	
		r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM
1	81	.98	2.1	—	—	.96	3.0	.98	2.1	—	—	.98	2.1	.97	2.6	.97	2.6	.98	2.1	.94	3.7	.83	6.2
2	119	—	—	.97	2.6	.95	3.4	—	—	.98	2.1	.95	3.4	.98	2.1	.96	3.0	.98	2.1	.96	3.0	—	—
3	171	—	—	.98	2.1	.95	3.4	—	—	.98	2.1	.97	2.6	.97	2.6	.96	3.0	.98	2.1	.97	2.6	—	—
4	147	—	—	.97	2.6	.95	3.4	—	—	.98	2.1	.96	3.0	.96	3.0	.94	3.7	.97	2.6	.96	3.0	—	—
5	140	—	—	.97	2.6	.95	3.4	—	—	.97	2.6	.93	4.0	.98	2.1	.93	4.0	.98	2.1	.96	3.0	—	—
6	147	—	—	.97	2.6	.95	3.4	—	—	.98	2.1	.88	5.2	.97	2.6	.95	3.4	.98	2.1	.95	3.4	—	—
7	131	—	—	.97	2.6	.94	3.7	—	—	.97	2.6	.89	5.0	.97	2.6	.95	3.4	.98	2.1	.95	3.4	—	—
8	106	—	—	.97	2.6	.95	3.4	—	—	.97	2.6	.95	3.4	.97	2.6	.94	3.7	.98	2.1	.94	3.7	—	—
9	104	—	—	.96	3.0	.94	3.7	—	—	.96	3.0	.87	5.4	.97	2.6	.95	3.4	.98	2.1	.94	3.7	—	—
10	91	—	—	.97	2.6	.96	3.0	—	—	.97	2.6	.96	3.0	.96	3.0	.95	3.4	.97	2.6	.95	3.4	—	—
11	89	—	—	.98	2.1	.96	3.0	—	—	.97	2.6	.96	3.0	.97	2.6	.95	3.4	.98	2.1	.96	3.0	—	—
12	75	—	—	.98	2.1	.96	3.0	—	—	.97	2.6	.97	2.6	.97	2.6	.96	3.0	.98	2.1	.95	3.4	—	—

Note. N = 1,401. Internal consistency estimates for composites were calculated using the reliability of linear combinations (Nunnally and Bernstein, 1994).  $SEM = SD \cdot \sqrt{1-r}$ , where SEM is the standard error of measurement, SD is the standard deviation of the standard score unit (15), and r is the reliability coefficient. **DDI (WRF)** = Dyslexia Diagnostic Index (Word Reading Fluency); **DDI (QRF)** = Dyslexia Diagnostic Index (Question Reading Fluency); **LPI** = Linguistic Processing Index; **RSI (WRF)** = Reading and Spelling Index (Word Reading Fluency); **RSI (QRF)** = Reading and Spelling Index (Question Reading Fluency); **SWA** = Sight Word Acquisition composite; **PK** = Phonics Knowledge composite; **BRS** = Basic Reading Skills composite; **DE** = Decoding Efficiency composite; **SP** = Spelling composite; **RF (WRF)** = Reading Fluency composite (Word Reading Fluency); **RF (QRF)** = Reading Fluency composite (Question Reading Fluency); **RC (SREI)** = Reading Comprehension Efficiency 1 composite (Silent Reading Efficiency Grades 1-5); **RC (SREZ)** = Reading Comprehension Efficiency 2 composite (Silent Reading Efficiency Grade 6-Adult); **PA** = Phonological Awareness composite; **RAN** = Rapid Automatized Naming composite; **AWM** = Auditory Working Memory composite; **OP** = Orthographic Processing composite; **VO** = Vocabulary composite; **RE** = Reasoning composite; **VR2** = Vocabulary and Reasoning 2 composite; **VR4** = Vocabulary and Reasoning 4 composite.

Table 5.4. TOD-C Internal Consistency Estimates and SEMs for Indexes and Composites by Grade: Child (continued)

Grade	n	Composites																							
		RF (QRF)		RC (SRE1)		RC (SRE2)		PA		RAN		AWM		OP		VO		RE		VR2		VR4			
		r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM		
1	81	—	—	—	—	—	.95	3.4	.99	1.5	.90	4.7	.90	4.7	.86	5.6	.89	5.0	.89	5.0	.89	5.0	.91	4.5	
2	119	.96	3.0	.98	2.1	—	.95	3.4	.99	1.5	.83	6.2	.81	6.5	.85	5.8	.91	4.5	.91	4.5	.91	4.5	.92	4.2	
3	171	.96	3.0	.98	2.1	—	.97	2.6	.99	1.5	.88	5.2	.85	5.8	.83	6.2	.92	4.2	.92	4.2	.90	4.7	.92	4.2	
4	147	.95	3.4	.98	2.1	—	.93	4.0	.99	1.5	.89	5.0	.85	5.8	.86	5.6	.92	4.3	.92	4.3	.90	4.7	.92	4.3	
5	140	.96	3.0	.98	2.1	—	.95	3.4	.99	1.5	.91	4.5	.84	6.0	.87	5.4	.92	4.3	.92	4.3	.91	4.5	.93	4.0	
6	147	.96	3.0	—	—	.99	.93	4.0	.99	1.5	.89	5.0	.85	5.8	.87	5.4	.93	4.0	.93	4.0	.92	4.2	.94	3.7	
7	131	.96	3.0	—	—	.99	.93	4.0	.99	1.5	.90	4.7	.84	6.0	.86	5.6	.93	4.0	.93	4.0	.91	4.5	.94	3.7	
8	106	.95	3.4	—	—	.99	.94	3.7	.99	1.5	.92	4.2	.87	5.4	.84	6.0	.91	4.5	.91	4.5	.92	4.4	.92	4.3	
9	104	.95	3.4	—	—	.99	.94	3.7	.99	1.5	.89	5.0	.81	6.5	.85	5.8	.91	4.5	.91	4.5	.90	4.7	.92	4.3	
10	91	.96	3.0	—	—	.99	.95	3.4	.99	1.5	.92	4.4	.85	5.8	.87	5.4	.92	4.3	.92	4.3	.91	4.5	.92	4.1	
11	89	.95	3.4	—	—	.99	.96	3.0	.99	1.5	.92	4.3	.85	5.8	.90	4.7	.94	3.7	.94	3.7	.92	4.3	.95	3.4	
12	75	.96	3.0	—	—	.99	.95	3.4	.99	1.5	.93	4.0	.90	4.7	.90	4.7	.93	4.0	.93	4.0	.92	4.3	.95	3.4	

Note. N = 1,401. Internal consistency estimates for composites were calculated using the reliability of linear combinations (Nunnally and Bernstein, 1994).  $SEM = SD \cdot \sqrt{1-r}$ , where SEM is the standard error of measurement, SD is the standard deviation of the standard score unit (15), and r is the reliability coefficient. **DDI (WRF)** = Dyslexia Diagnostic Index (Word Reading Fluency); **DDI (QRF)** = Dyslexia Diagnostic Index (Question Reading Fluency); **LPI** = Linguistic Processing Index; **RSI (WRF)** = Reading and Spelling Index (Word Reading Fluency); **RSI (QRF)** = Reading and Spelling Index (Question Reading Fluency); **SWA** = Sight Word Acquisition composite; **PK** = Phonics Knowledge composite; **BRS** = Basic Reading Skills composite; **DE** = Decoding Efficiency composite; **SP** = Spelling composite; **RF (WRF)** = Reading Fluency composite (Word Reading Fluency); **RF (QRF)** = Reading Fluency composite (Question Reading Fluency); **RC (SRE1)** = Reading Comprehension Efficiency 1 composite (Silent Reading Efficiency Grades 1–5); **RC (SRE2)** = Reading Comprehension Efficiency 2 composite (Silent Reading Efficiency Grade 6–Adult); **PA** = Phonological Awareness composite; **RAN** = Rapid Automated Naming composite; **AWM** = Auditory Working Memory composite; **OP** = Orthographic Processing composite; **VO** = Vocabulary composite; **RE** = Reasoning composite; **VR2** = Reasoning and Reasoning 2 composite; **VR4** = Vocabulary and Reasoning 4 composite.



Table 5.5. TOD-C Internal Consistency Estimates and SEMs for Indexes and Composites by Age Range: Adult

Age (years)	n	Indexes						Composites											
		DDI (QRF)		LPI		RSI (QRF)		SWA		PK		BRS		DE		SP		RF (QRF)	
		r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM
18-23	113	.92	4.2	.96	3.0	.97	2.6	.96	3.0	.95	3.4	.95	3.4	.98	2.1	.96	3.0	.96	3.0
24-39	64	.93	4.0	.96	3.0	.97	2.6	.95	3.4	.96	3.0	.93	4.0	.96	3.0	.93	4.0	.96	3.0
40-49	40	.92	4.4	.96	3.0	.95	3.4	.95	3.4	.96	3.0	.91	4.5	.95	3.4	.93	4.0	.97	2.6
50-59	54	.96	3.0	.97	2.6	.98	2.1	.95	3.4	.98	2.1	.93	4.0	.98	2.1	.96	3.0	.96	3.0
60-69	37	.96	3.0	.98	2.1	.98	2.1	.97	2.6	.97	2.6	.97	2.6	.98	2.1	.95	3.4	.96	3.0
70-89	39	.96	3.0	.97	2.6	.98	2.1	.91	4.5	.97	2.6	.97	2.6	.97	2.6	.97	2.6	.96	3.0

Age (years)	n	Composites						Composites						Composites					
		RC (SRE2)		PA		RAN		AWM		OP		VO		RE		VR2		VR4	
		r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM	r	SEM
18-23	113	.93	4.0	.94	3.7	.99	1.5	.83	6.2	.93	4.0	.93	4.0	.90	4.7	.93	4.0	.92	4.2
24-39	64	.99	1.5	.95	3.4	.99	1.5	.84	6.0	.91	4.5	.94	3.7	.90	4.7	.93	4.0	.93	4.0
40-49	40	.99	1.5	.95	3.4	.99	1.5	.85	5.8	.92	4.2	.93	4.0	.90	4.7	.94	3.7	.91	4.5
50-59	54	.99	1.5	.97	2.6	.99	1.5	.88	5.2	.93	4.0	.94	3.7	.94	3.7	.95	3.4	.95	3.4
60-69	37	.99	1.5	.98	2.1	.99	1.5	.90	4.7	.94	3.7	.94	3.7	.94	3.7	.95	3.4	.96	3.0
70-89	39	.93	4.0	.98	2.1	.99	1.5	.85	5.8	.92	4.2	.94	3.7	.97	2.6	.94	3.7	.97	2.6

Note: N = 347. Internal consistency estimates for composites were calculated using the reliability of linear combinations (Nunnally and Bernstein, 1994).  $SEM = SD \sqrt{1 - r}$ , where SEM is the standard error of measurement, SD is the standard deviation of the standard score unit (15), and r is the reliability coefficient. **DDI (QRF)** = Dyslexia Diagnostic Index (Question Reading Fluency); **LPI** = Linguistic Processing Index; **RSI (QRF)** = Reading and Spelling Index (Question Reading Fluency); **SWA** = Sight Word Acquisition composite; **PK** = Phonics Knowledge composite; **BRS** = Basic Reading Skills composite; **DE** = Decoding Efficiency composite; **SP** = Spelling composite; **RF (QRF)** = Reading Fluency composite (Question Reading Fluency); **RC (SRE2)** = Reading Comprehension Efficiency 2 composite (Silent Reading Efficiency Grade 6-Adult); **PA** = Phonological Awareness composite; **RAN** = Rapid Automated Naming composite; **AWM** = Auditory Working Memory composite; **OP** = Orthographic Processing composite; **VO** = Vocabulary composite; **RE** = Reasoning composite; **VR2** = Vocabulary and Reasoning 2 composite; **VR4** = Vocabulary and Reasoning 4 composite.



**Table 5.6.** TOD-C Internal Consistency Estimates and SEMs for the Rating Scale Standardization Sample by Age Range

Age (years)	Self-Rating			Parent/Caregiver Rating			Teacher Rating		
	<i>n</i>	<i>r</i>	<i>SEM</i>	<i>n</i>	<i>r</i>	<i>SEM</i>	<i>n</i>	<i>r</i>	<i>SEM</i>
6–7	65	.94	2.4	68	.94	2.4	31	.97	1.7
8–9	210	.91	3.0	186	.94	2.4	93	.96	2.0
10–11	186	.94	2.4	187	.95	2.2	95	.97	1.7
12–13	188	.94	2.4	192	.96	2.0	88	.97	1.7
14–15	147	.94	2.4	130	.93	2.6	42	.95	2.2
16–18	151	.94	2.4	123	.96	2.0	44	.97	1.7
19–23	75	.93	2.6	—	—	—	—	—	—
24–49	85	.94	2.4	—	—	—	—	—	—
50–89	94	.95	2.2	—	—	—	—	—	—

Note. *N* = 1,452. Parent/Caregiver and Teacher Ratings are for individuals Grades 1–12. Internal consistency estimates were calculated using Cronbach's alpha.  $SEM = SD \sqrt{1 - r}$ , where *SEM* is the standard error of measurement, *SD* is the standard deviation of the *T*-score unit (10), and *r* is the reliability coefficient.

**Table 5.7.** TOD-E Internal Consistency Estimates and SEMs for Tests by Age

Age (years) <sup>a</sup>	<i>n</i>	SPW-E		RHY-E		ERNL-E <sup>b</sup>		LSW-E		ESEG-E		LSK-E	
		<i>r</i>	<i>SEM</i>	<i>r</i>	<i>SEM</i>	<i>r</i>	<i>SEM</i>	<i>r</i>	<i>SEM</i>	<i>r</i>	<i>SEM</i>	<i>r</i>	<i>SEM</i>
5	72	.92	4.2	.96	3.0	.99	1.5	.93	4.0	.96	3.0	.95	3.4
6	122	.93	4.0	.94	3.7	.99	1.5	.97	2.6	.96	3.0	.96	3.0
7	104	.90	4.7	.93	4.0	.97	2.6	.96	3.0	.83	6.2	.91	4.5
8–9:3	44	.87	5.4	.92	4.2	.99	1.5	.92	4.2	.92	4.2	.93	4.0

Note. *N* = 342. Internal consistency estimates for timed tests were calculated using Rasch-based reliability; all others were based on the split-half method.  $SEM = SD \sqrt{1 - r}$ , where *SEM* is the standard error of measurement, *SD* is the standard deviation of the standard score unit (15), and *r* is the reliability coefficient. SPW-E = Sounds and Pseudowords; RHY-E = Rhyming; ERNL-E = Early Rapid Number and Letter Naming; LSW-E = Letter and Sight Word Recognition; ESEG-E = Early Segmenting; LSK-E = Letter and Sound Knowledge.

<sup>a</sup>8-year normative group extends through age 9 years, 3 months.

<sup>b</sup>Timed test.

**Table 5.8.** TOD-E Internal Consistency Estimates and SEMs for Indexes and Composites by Grade

Grade	<i>n</i>	Indexes										Composites							
		EDDI (WRF)		EDDI (QRF)		ELPI		ERSI (WRF)		ERSI (QRF)		ESWA		EPK		EBRS		EPA	
		<i>r</i>	<i>SEM</i>	<i>r</i>	<i>SEM</i>	<i>r</i>	<i>SEM</i>	<i>r</i>	<i>SEM</i>	<i>r</i>	<i>SEM</i>	<i>r</i>	<i>SEM</i>	<i>r</i>	<i>SEM</i>	<i>r</i>	<i>SEM</i>	<i>r</i>	<i>SEM</i>
K	122	.98	2.1	.98	2.1	.98	2.1	.97	2.6	.97	2.6	.93	4.0	.96	3.0	.97	2.6	.96	3.0
1	118	.98	2.1	.98	2.1	.97	2.6	.96	3.0	.96	3.0	.95	3.4	.93	4.0	.97	2.6	.96	3.0
2	102	.97	2.6	.97	2.6	.95	3.4	.96	3.0	.96	3.0	.92	4.2	.93	4.0	.96	3.0	.94	3.7

Note. *N* = 342. Internal consistency estimates for composites were calculated using the reliability of linear combinations (Nunnally and Bernstein, 1994).  $SEM = SD \sqrt{1 - r}$ , where *SEM* is the standard error of measurement, *SD* is the standard deviation of the standard score unit (15), and *r* is the reliability coefficient. EDDI (WRF) = Early Dyslexia Diagnostic Index (Word Reading Fluency); EDDI (QRF) = Early Dyslexia Diagnostic Index (Question Reading Fluency); ELPI = Early Linguistic Processing Index; ERSI (WRF) = Early Reading and Spelling Index (Word Reading Fluency); ERSI (QRF) = Early Reading and Spelling Index (Question Reading Fluency); ESWA = Early Sight Word Acquisition composite; EPK = Early Phonics Knowledge composite; EBRS = Early Basic Reading Skills composite; EPA = Early Phonological Awareness composite.

**Table 5.9.** TOD-E Internal Consistency Estimates and SEMs for the Rating Scale Standardization Sample by Age Range

Age (years) <sup>a</sup>	Parent/Caregiver Rating			Teacher Rating		
	<i>n</i>	<i>r</i>	<i>SEM</i>	<i>n</i>	<i>r</i>	<i>SEM</i>
5–6	100	.95	2.2	91	.96	2.0
7–9:3	74	.96	2.0	72	.97	1.7

Note. *N* = 211. Internal consistency estimates were calculated using Cronbach's alpha.  $SEM = SD \sqrt{1 - r}$ , where *SEM* is the standard error of measurement, *SD* is the standard deviation of the *T*-score unit (10), and *r* is the reliability coefficient.

<sup>a</sup>8-year normative group extends through age 9 years, 3 months.

## Test–Retest Reliability

*Test–retest reliability*, also known as *temporal stability*, defines the extent to which an examinee's test scores remain the same over time, assuming the underlying ability does not change. It is estimated for each test by administering the same test to the same individual on two separate occasions, typically only two or three weeks apart, and then calculating the correlation coefficient between the two sets of scores. Over this brief interval, test scores are not expected to change appreciably due to development of the underlying abilities. However, scores may change as a result of random variations in performance, or because of learning due to repeated exposure to the same test stimuli.

A total of 90 individuals participated in test–retest studies across the TOD-S, TOD-C, and TOD-E. The TOD-S sample included 81 individuals ranging in age from 5 to 54 years ( $M = 14.37$  years,  $SD = 10.93$ ), split evenly between males and females, and was 65% Hispanic and 37% White. In terms of head-of-household education level, 52% had a high school diploma or lower, and 48% had at least some college. Sixty-one individuals were in the TOD-C retest sample, ranging in age from 8 to 54 years ( $M = 17.43$  years,  $SD = 11.71$ ), and 30 were in the TOD-E sample, ranging in age from 5 to 8 years ( $M = 6.38$  years,  $SD = .92$ ). Because the individuals who took the TOD-C or TOD-E also took the TOD-S, the sample compositions are demographically similar.

Tables 5.10 to 5.12 present the results of the test–retest reliability studies for the TOD-S, TOD-C, and TOD-E, respectively. Across all three samples, the reliability coefficients for tests, indexes, and composites range from .70 to .97, with a median of .88. These coefficients are satisfactory for tests of developing abilities. To illustrate temporal stability in another way, Tables 5.10 to 5.12 also show the means and

standard deviations for the Time 1 and Time 2 standard scores, as well as the effect size of the difference between the means. The effect size was calculated as the difference between the mean standard scores of the two testing occasions, divided by the pooled standard deviation. By this method, an effect size of 0.2 is considered small, 0.5 is considered medium, and 0.8 is considered large (Cohen, 1992). The effect sizes range from 0.01 to 0.48 (median = 0.18) across the TOD test, index, and composite scores. Although most effect sizes are considered small, indicating negligible change from Time 1 to Time 2 in the average performance of the test–retest sample, a few approach the medium range, which is not unexpected due to the higher likelihood of practice effects with certain tests. Taken as a whole, these results support the stability of TOD scores over time and also indicate the importance of delaying a second administration for a longer period of time (at least three months) to avoid any practice effects.

## Standard Error of Measurement (SEM) and Confidence Intervals

The *standard error of measurement (SEM)* is calculated from the reliability coefficient and is used to create a confidence interval, i.e., a range of scores that contains the examinee's "true score" within a given probability, e.g., .90. The true score refers to the hypothetical mean that would be obtained from repeated testing minus the effects of practice, fatigue, and other sources of error. The *SEM* is calculated using the formula  $SEM = SD \sqrt{1 - r}$ , where *SD* is the standard deviation of the scale and *r* is the reliability coefficient of the scale. *SEM* values are displayed in Tables 5.1 to 5.9 next to their respective internal consistency reliability coefficients, and in Tables 5.10 to 5.12 next to their respective test–retest reliability coefficients.

*SEM* values can be converted into confidence intervals that give a range of probable values for the true score. For example, the 90% confidence interval represents the range of scores around the observed score that has a 90% probability of containing the true score; the 95% confidence interval represents the range of scores around the observed score that has a 95% probability of containing the true score. The confidence values for each test are provided in the scoring tables in the appendices. They are expressed in standard score units (*T*-scores for the Rating Scales) and rounded to the nearest whole number. In most cases, *SEMs* of internal consistency reliability coefficients are used as the basis for confidence intervals; however, in the case of speeded tests, test-retest reliability coefficients are used. Rasch-based internal consistency estimates are high (as described previously), and therefore *SEMs* based on these internal consistency estimates are too narrow to be used in clinical practice. *SEMs* from Rasch-based internal consistency reliability estimates for tests with item sets are an appropriate basis for confidence values. Confidence intervals are based on reliability estimates for the whole sample.

Chapter 2 describes the procedure for using confidence values to determine confidence intervals, and Chapter 3 presents interpretation of the confidence intervals.

## Rating Scale Cross-Form Consistency

*Cross-form consistency* refers to studies in which respondents rate an individual on two different forms (e.g., parent and teacher ratings). The cross-form ratings may vary because the two respondents are observing the individual being rated in varying environments and at different times. Similarly, an individual may have a different perspective on themselves than do the adults around them.

Cross-form analyses of the TOD Rating Scales for individuals who had at least two ratings were conducted separately for the TOD-C child sample and the TOD-E sample. Across all analyses, strong correlations were found between raters: TOD-E Parent/Caregiver and Teacher Rating forms ( $r = .75, n = 85$ ); TOD-C Parent/Caregiver and Self-Rating forms ( $r = .63, n = 880$ ); TOD-C Parent/Caregiver and Teacher Rating forms ( $r = .73, n = 344$ ); and TOD-C Self- and Teacher Rating forms ( $r = .60, n = 391$ ). These results indicate that different raters' responses contribute unique variance, though most scores are likely to be similar. In addition, the differences between ratings by multiple respondents provide more breadth of information about the individual being rated. Mean *T*-score differences between raters were approximately one half of a standard deviation for all comparisons (TOD-C Parent/Caregiver and Teacher Rating difference mean = 4.5; TOD-C Parent/Caregiver and Self-Rating difference mean = 5.28; TOD-C Teacher and Self-Rating difference mean = 5.70; TOD-E Parent/Caregiver and Teacher Rating difference mean = 4.60).

**Table 5.10.** TOD-S Test–Retest Reliability: Descriptive Statistics, Effect Sizes, Corrected Correlations, and *SEMs*

Test/Index	Time 1		Time 2		Effect size	<i>r</i>	Corrected <i>r</i> <sup>a</sup>	<i>SEM</i>
	Mean	<i>SD</i>	Mean	<i>SD</i>				
<b>Test</b>								
Picture Vocabulary	96.85	16.65	99.23	17.94	0.14	.78	.75	7.48
Letter and Word Choice	95.20	15.20	94.35	15.35	0.06	.77	.76	7.29
Word Reading Fluency	98.30	14.09	101.35	14.73	0.22	.93	.94	3.81
Question Reading Fluency	97.10	13.58	98.85	14.70	0.13	.95	.96	3.15
<b>Index</b>								
Dyslexia Risk Index (WRF)	96.80	17.47	101.90	16.06	0.29	.95	.94	3.82
Dyslexia Risk Index (QRF)	95.56	13.77	94.98	14.56	0.04	.88	.90	4.80

Note.  $N = 81$ . Means, *SDs* expressed in standard score units ( $M = 100, SD = 15$ ). Effect size (Cohen's  $d$ ) = Time 2 mean minus Time 1 mean, divided by pooled *SD*, where pooled *SD* is  $\sqrt{[(\text{Time } 1 \ n) \times (\text{Time } 1 \ SD^2) + (\text{Time } 2 \ n) \times (\text{Time } 2 \ SD^2)] / (\text{Time } 1 \ n + \text{Time } 2 \ n)}$ . WRF = Word Reading Fluency; QRF = Question Reading Fluency.

<sup>a</sup>The reliability coefficient ( $r$ ) was corrected for variability of normative group ( $SD = 15$ ) based on standard deviation obtained at Time 1, using Guilford's (1954) formula.

**Table 5.11.** TOD-C Test–Retest Reliability:  
Descriptive Statistics, Effect Sizes, Corrected Correlations, and SEMs

Test/Index/Composite	Time 1		Time 2		Effect size	<i>r</i>	Corrected <i>r</i> <sup>a</sup>	SEM
	Mean	SD	Mean	SD				
<b>Test</b>								
Phonological Manipulation	97.85	15.58	99.69	15.54	0.12	.86	.85	5.82
Irregular Word Spelling	96.39	13.46	98.20	14.95	0.13	.89	.91	4.53
Rapid Letter Naming	100.46	14.43	100.31	14.91	0.01	.83	.84	6.00
Pseudoword Reading	99.49	13.98	100.61	12.82	0.08	.71	.74	7.72
Word Pattern Choice	98.92	14.45	105.41	15.35	0.45	.81	.82	6.34
Word Memory	94.38	14.29	96.38	15.47	0.14	.77	.78	7.00
Picture Analogies	93.49	14.02	97.89	13.69	0.31	.72	.74	7.63
Irregular Word Reading	94.31	14.22	97.52	13.63	0.23	.70	.72	7.93
Oral Reading Efficiency	97.61	15.35	101.38	15.01	0.25	.89	.89	5.07
Blending	100.54	15.45	104.20	14.54	0.24	.82	.81	6.47
Segmenting	96.90	17.22	100.03	14.56	0.18	.83	.79	6.94
Regular Word Spelling	97.56	14.18	99.75	14.16	0.15	.93	.94	3.74
Silent Reading Efficiency Grades 1–5	89.80	16.80	95.67	21.72	0.35	.94	.93	4.10
Silent Reading Efficiency Grade 6–Adult	96.98	11.79	101.15	12.45	0.35	.80	.86	5.63
Rapid Number and Letter Naming	100.13	13.12	101.02	15.39	0.07	.80	.84	6.06
Letter Memory	95.26	14.58	97.34	17.12	0.14	.79	.80	6.77
Rapid Pseudoword Reading	99.39	13.77	102.38	15.42	0.22	.79	.81	6.53
Rapid Irregular Word Reading	97.37	10.08	99.20	10.90	0.18	.72	.84	6.08
Symbol to Sound Learning	84.67	14.91	85.79	15.26	0.07	.73	.73	7.82
Listening Vocabulary	94.89	14.26	95.10	13.96	0.02	.88	.89	5.05
Geometric Analogies	95.52	12.96	97.33	15.12	0.14	.85	.88	5.22

Note. *N* = 61 (Silent Reading Efficiency Grade 6–Adult *n* = 46). Means, *SD*s expressed in standard score units (*M* = 100, *SD* = 15). Effect size (Cohen's *d*) = Time 2 mean minus Time 1 mean, divided by pooled *SD*, where pooled *SD* is  $\sqrt{[(\text{Time } 1n) \times (\text{Time } 1SD)^2 + (\text{Time } 2n) \times (\text{Time } 2SD)^2] / (\text{Time } 1n + \text{Time } 2n)}$ . Composites including WRF-S and/or SRE1-C [i.e., DDI (WRF), RSI (WRF), RF (WRF), RC (SRE1)] were not included in the analyses due to small sample sizes. QRF = Question Reading Fluency; SRE2 = Silent Reading Efficiency Grade 6–Adult; WRF-S = Word Reading Fluency; SRE1-C = Silent Reading Efficiency Grades 1–5; DDI (WRF) = Dyslexia Diagnostic Index (Word Reading Fluency); RSI (WRF) = Reading and Spelling Index (Word Reading Fluency); RF (WRF) = Reading Fluency composite (Word Reading Fluency); RC (SRE1) = Reading Comprehension Efficiency 1 composite (Silent Reading Efficiency Grades 1–5).

<sup>a</sup>The reliability coefficient (*r*) was corrected for variability of normative group (*SD* = 15) based on standard deviation obtained at Time 1, using Guilford's (1954) formula.

**Table 5.11.** TOD-C Test–Retest Reliability:  
Descriptive Statistics, Effect Sizes, Corrected Correlations, and SEMs (continued)

Test/Index/Composite	Time 1		Time 2		Effect size	<i>r</i>	Corrected <i>r</i> <sup>a</sup>	SEM
	Mean	SD	Mean	SD				
<b>Index</b>								
Dyslexia Diagnostic Index (QRF)	98.13	15.76	95.67	14.93	0.16	.93	.92	4.12
Linguistic Processing Index	100.39	15.84	96.77	15.22	0.23	.89	.88	5.11
Reading and Spelling Index (QRF)	96.13	14.38	95.49	13.65	0.04	.91	.92	4.34
<b>Composite</b>								
Sight Word Acquisition composite	98.34	11.65	95.77	10.72	0.22	.81	.87	5.33
Phonics Knowledge composite	101.70	14.83	99.15	14.27	0.17	.79	.80	6.78
Basic Reading Skills composite	98.80	13.46	96.38	12.85	0.18	.76	.79	6.80
Decoding Efficiency composite	101.02	13.49	98.33	12.11	0.20	.79	.82	6.33
Spelling composite	98.93	14.75	96.92	13.77	0.14	.95	.95	3.24
Reading Fluency (QRF) composite	99.75	15.17	95.87	14.53	0.26	.95	.94	3.56
Reading Comprehension Efficiency 2 (SRE2) composite	99.93	13.42	96.26	12.24	0.27	.90	.92	4.24
Phonological Awareness composite	101.25	14.96	97.82	17.09	0.23	.88	.88	5.10
Rapid Automatized Naming composite	101.08	15.35	100.74	13.58	0.02	.84	.84	6.09
Auditory Working Memory composite	96.30	18.01	93.93	15.88	0.13	.84	.79	6.85
Orthographic Processing composite	98.66	15.01	96.79	14.34	0.12	.81	.81	6.58
Vocabulary composite	94.31	15.26	93.49	15.92	0.05	.89	.89	5.07
Reasoning composite	97.35	15.09	93.76	13.87	0.24	.86	.86	5.66
Vocabulary and Reasoning 2 composite	94.92	14.70	92.02	15.79	0.20	.85	.85	5.76
Vocabulary and Reasoning 4 composite	95.40	15.97	92.60	15.41	0.18	.93	.93	4.08

Note. *N* = 61 (Silent Reading Efficiency Grade 6–Adult *n* = 46). Means, *SD*s expressed in standard score units (*M* = 100, *SD* = 15). Effect size (Cohen's *d*) = Time 2 mean minus Time 1 mean, divided by pooled *SD*, where pooled *SD* is  $\sqrt{[(\text{Time } 1n) \times (\text{Time } 1SD^2) + (\text{Time } 2n) \times (\text{Time } 2SD^2)] / (\text{Time } 1n + \text{Time } 2n)}$ . Composites including WRF-S and/or SRE1-C [i.e., DDI (WRF), RSI (WRF), RF (WRF), RC (SRE1)] were not included in the analyses due to small sample sizes. **QRF** = Question Reading Fluency; **SRE2** = Silent Reading Efficiency Grade 6–Adult; **WRF-S** = Word Reading Fluency; **SRE1-C** = Silent Reading Efficiency Grades 1–5; **DDI (WRF)** = Dyslexia Diagnostic Index (Word Reading Fluency); **RSI (WRF)** = Reading and Spelling Index (Word Reading Fluency); **RF (WRF)** = Reading Fluency composite (Word Reading Fluency); **RC (SRE1)** = Reading Comprehension Efficiency 1 composite (Silent Reading Efficiency Grades 1–5).

<sup>a</sup>The reliability coefficient (*r*) was corrected for variability of normative group (*SD* = 15) based on standard deviation obtained at Time 1, using Guilford's (1954) formula.

**Table 5.12.** TOD-E Test–Retest Reliability: Descriptive Statistics, Effect Sizes, Corrected Correlations, and SEMs

Test/Index/Composite	Time 1		Time 2		Effect size	<i>r</i>	Corrected <i>r</i> <sup>a</sup>	SEM
	Mean	SD	Mean	SD				
<b>Test</b>								
Sounds and Pseudowords	105.41	14.06	106.97	17.36	0.11	.92	.93	4.03
Rhyming	103.07	14.59	105.55	16.51	0.17	.86	.87	5.41
Early Rapid Number and Letter Naming	103.62	12.64	105.86	12.79	0.18	.82	.86	5.64
Letter and Sight Word Recognition	99.86	12.77	102.66	12.44	0.22	.95	.96	2.94
Early Segmenting	102.20	11.20	107.56	10.58	0.48	.88	.92	4.13
Letter and Sound Knowledge	99.41	15.76	103.93	14.86	0.29	.89	.88	5.12
<b>Index</b>								
Early Dyslexia Diagnostic Index (WRF)	96.88	18.42	104.13	17.62	0.39	.98	.97	2.71
Early Linguistic Processing Index	103.60	13.75	108.28	14.11	0.34	.91	.92	4.19
Early Reading and Spelling Index (WRF)	97.32	18.73	104.11	17.33	0.36	.97	.96	3.10
<b>Composite</b>								
Early Sight Word Acquisition composite	96.50	16.79	102.57	14.55	0.36	.95	.93	3.83
Early Phonics Knowledge composite	103.07	16.31	106.70	17.88	0.22	.95	.94	3.69
Early Basic Reading Skills composite	100.07	16.27	104.33	15.45	0.26	.94	.92	4.11
Early Phonological Awareness composite	103.20	14.42	108.48	14.09	0.37	.91	.91	4.40

Note. *N* = 30; some tests/composites have fewer cases. Means, *SD*s expressed in standard score units (*M* = 100, *SD* = 15). Effect size (Cohen's *d*) = Time 2 mean minus Time 1 mean, divided by pooled *SD*, where pooled *SD* is  $\sqrt{[(\text{Time } 1n) \times (\text{Time } 1SD^2) + (\text{Time } 2n) \times (\text{Time } 2SD^2)] / (\text{Time } 1n + \text{Time } 2n)}$ . Composites including QRF-S [i.e., EDDI (QRF), ERSI (QRF)] were not included in the analyses due to small sample sizes. **WRF** = Word Reading Fluency; **QRF-S** = Question Reading Fluency; **EDDI (QRF)** = Early Dyslexia Diagnostic Index (Question Reading Fluency); **ERSI (QRF)** = Early Reading and Spelling Index (Question Reading Fluency).

<sup>a</sup>The reliability coefficient (*r*) was corrected for variability of normative group (*SD* = 15) based on standard deviation obtained at Time 1, using Guilford's (1954) formula.



# Validity

---

At the most fundamental level, tests and rating scales are considered valid if they measure what they are supposed to measure. Validation evidence must be presented for a test's well-defined purposes, under specified conditions, and for the populations with which it is intended to be used. This section presents evidence addressing the TOD's content-description validity, construct validity, convergent validity, validity based on detection of skill weaknesses, clinical-groups validity, and predictive validity.

## Content-Description Validity

According to Anastasi and Urbina (1997), *content-description validity* requires “the systematic examination of the test content to determine whether it covers a representative sample of the behavior domain to be measured” (p. 115). Figure 1.1 in Chapter 1 shows the constructs assessed by the TOD and the tests developed to operationalize those constructs.

TOD test items were created based on theoretical fit and review of the literature. Test items were constructed to assess the pattern of abilities characterizing dyslexia as described in the research literature.

To ensure content validity of the TOD Rating Scales, the literature describing characteristics of dyslexia and its underlying etiology was reviewed, along with other related instruments. Based on these sources, items were created to elicit the relevant background/history associated with dyslexia and its most salient characterizations (e.g., Kilpatrick, 2015; Mather & Wendling, in press; Pennington et al., 2019): motivation for reading, general reasoning, verbal comprehension, orthographic processing, phonological awareness, rapid automatized naming, memory, basic reading skills, reading fluency, reading comprehension, and spelling. Each of the Rating Scales contains several Yes or No questions related to family history, history of reading support, grade retention, and previous diagnoses, followed by a set of Likert-type items with responses ranging from Strongly Disagree (1) to Strongly Agree (4); the higher the score on the Rating Scales, the greater the dyslexia risk. Gathering information regarding family history

is especially important given research indicating that having relatives with reading difficulty is a strong risk factor for dyslexia (Hamilton & Hayiou-Thomas, 2022; Lasnick et al., 2022; Snowling et al., 2019).

## Construct Validity

*Construct validity* is defined as the extent to which a test (or tests) accurately assesses a theoretical construct of interest and is determined by several sources of evidence. First, because the TOD was developed to assess reading ability and to be sensitive to the reading limitations associated with dyslexia, it contains a number of tests that should be correlated with reading skill development. Consequently, increases in TOD test raw scores should be related to chronological age- or grade-level progression. TOD test scores should also correlate more strongly with other measures of reading ability and related constructs than with theoretically unrelated constructs (both within the TOD and when compared with other assessments). Additionally, the factor structure of the TOD should represent the theoretical constructs of dyslexia described in the literature. Finally, TOD scores should differentiate between examinees known to have reading deficits consistent with dyslexia and those who do not.

## Developmental Progression

The constructs measured by the TOD display different developmental patterns that can provide additional validity support for the tests. All abilities measured by the TOD should show developmental variability. Thus, prior to creating standard scores, TOD raw score means were examined to ensure that they fit with the expectation of skill growth specific to the construct in question. All of the skills measured by the TOD should show early rapid growth that tapers off at different ages. For example, although vocabulary knowledge increases throughout the life span, it grows quickly beginning at age 3 and slows down around age 12 for most individuals (Byrnes, 2021). Considerable research indicates the developmental trajectories that different skills should take. The following sections illustrate the ways in which the TOD tests conform to expectations.



Tests that reach a ceiling during the middle school years are measures of skills that reach mastery and then do not continue to improve. These skills require a shorter growth period to reach proficiency. For example, phonological awareness skills follow this pattern. Once children have mastered how to rhyme words, their ability does not continue to improve. Other phonological skills mastered at younger ages include blending, segmenting, manipulating sounds, and knowledge of sound–symbol correspondences. In a school setting, phonological awareness skills and sound–symbol associations are taught in the early elementary grades. Most students will have mastered these skills by the middle school years. When a particular skill reaches maximum growth as measured by a particular test and there is no further development of that skill, a ceiling is reached, i.e., the test cannot discriminate skill acquisition any further. Most ability tests reflect this phenomenon (e.g., see Bracken & McCallum, 2016; McGrew et al., 2014; Wechsler, 2014).

Other tests reach a ceiling during high school years. Typically, these tests require application of previously learned skills. For example, in addition to

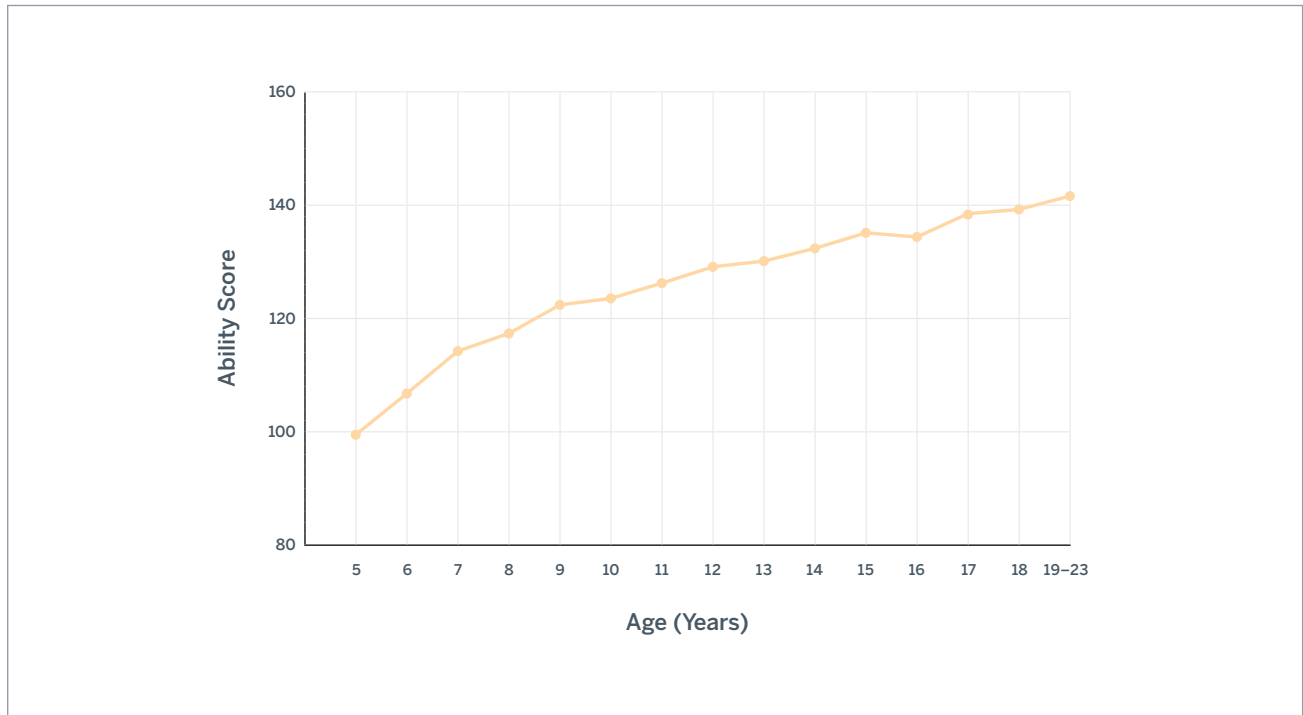
knowing letter and number names, rapid naming tasks require automaticity. Applying phonological skills and knowledge of common English spelling patterns is required when reading and spelling both pseudowords and irregular words. Reading fluency requires both intact and automatic word recognition. These types of tasks have a steeper growth curve than early literacy skills, so there is more room to measure proficiency.

Still other tests continue to show an increase in skills beyond high school. These are tests of linguistic processing skills (such as letter memory) or acquired knowledge (such as vocabulary). Like those skills that reach an earlier ceiling, these also show rapid growth in the early years that slows during high school. However, rather than reach a ceiling, the skills continue to grow until middle age, assuming a supportive environment. Analysis of these developmental differences helps explain the ceiling effects evident on some TOD tests.

## TOD-S

All TOD-S tests demonstrate rapid early growth and then a slower but consistently increasing trajectory through high school and beyond. The continued skill differentiation among the tests demonstrates their utility as a screener across the life span. All TOD-S test raw scores continue to grow beyond high school.

Figure 5.1 provides an illustration of this increase in means, using Picture Vocabulary (1S) as an example. This figure, as well as the others in this section, are cut off at age 23 because it represents the plateau of development in early adulthood. The Word Reading Fluency test (3Sa) covers only three years but shows consistent growth across the ages of 5–7 years.



**Figure 5.1.** Increase in Means Example: Picture Vocabulary (Test 1S)

**TOD-C**

All TOD-C tests demonstrate rapid early growth and then a slower but consistently increasing trajectory through at least middle school. As described above, it is expected for individuals to obtain mastery in

some skills earlier than others. Table 5.13 identifies which TOD tests reach a ceiling at middle school, high school, and beyond. Figures 5.2 to 5.4 illustrate examples of these three ceiling categories, showing the increase in means through age 23, which represents the plateau of development in early adulthood.

**Table 5.13.** TOD-C Test Ceilings

Middle school	High school	Beyond high school
<b>Phonological Awareness</b>	<b>Auditory Working Memory</b>	<b>Spelling</b>
Phonological Manipulation	Word Memory	Irregular Word Spelling
Blending	Letter Memory	Regular Word Spelling
Segmenting	<b>Rapid Automatized Naming</b>	<b>Vocabulary</b>
<b>Visual-Verbal Paired-Associate Learning</b>	Rapid Letter Naming	Listening Vocabulary
Symbol to Sound Learning	Rapid Number and Letter Naming	<b>Reasoning</b>
	<b>Orthographic Processing</b>	Picture Analogies
	Word Pattern Choice	Geometric Analogies
	<b>Phonics Knowledge</b>	<b>Word Reading</b>
	Pseudoword Reading	Irregular Word Reading
	Rapid Pseudoword Reading	Rapid Irregular Word Reading
	<b>Reading Fluency</b>	<b>Reading Comprehension Fluency</b>
	Oral Reading Efficiency	Silent Reading Efficiency



**Figure 5.2.** Ceiling in Middle School Example: Blending (Test 13C)

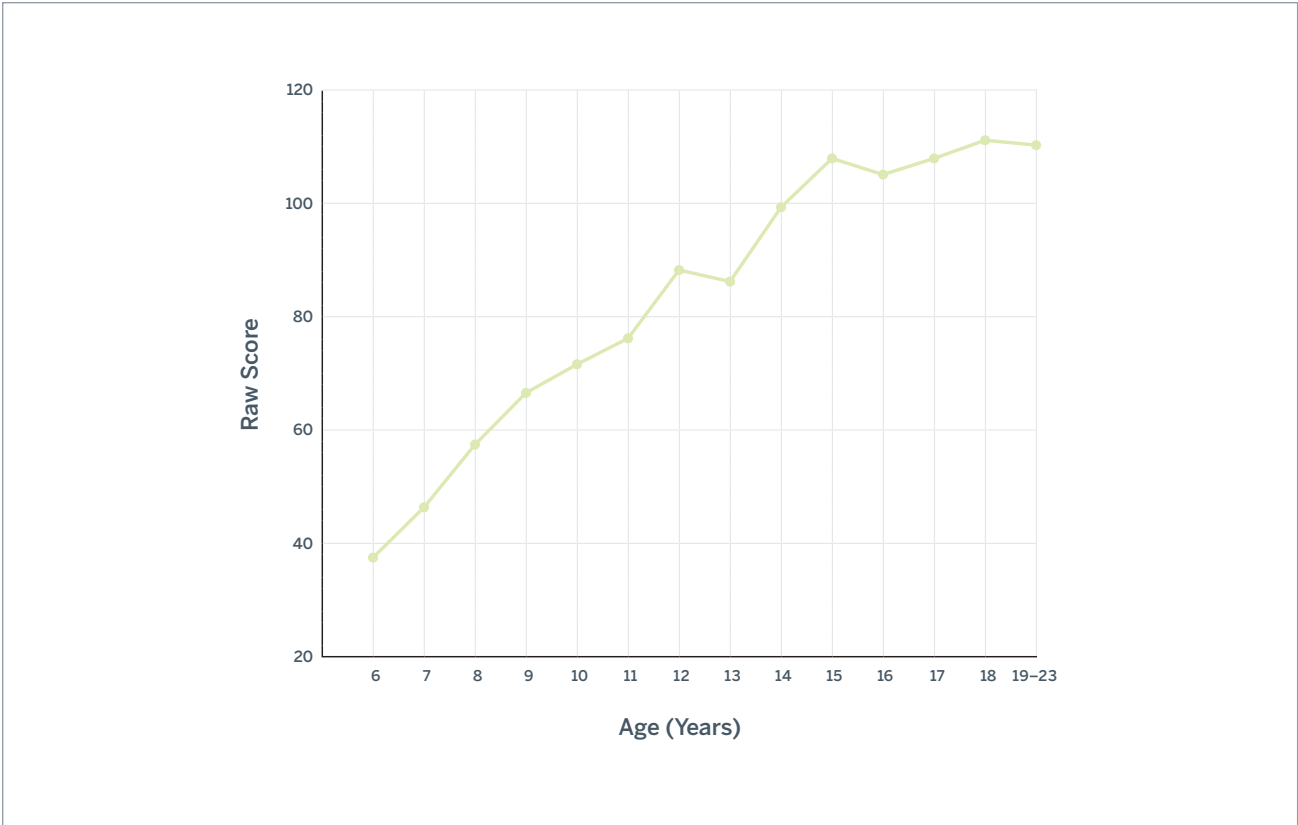


Figure 5.3. Ceiling in High School Example: Rapid Letter Naming (Test 6C)

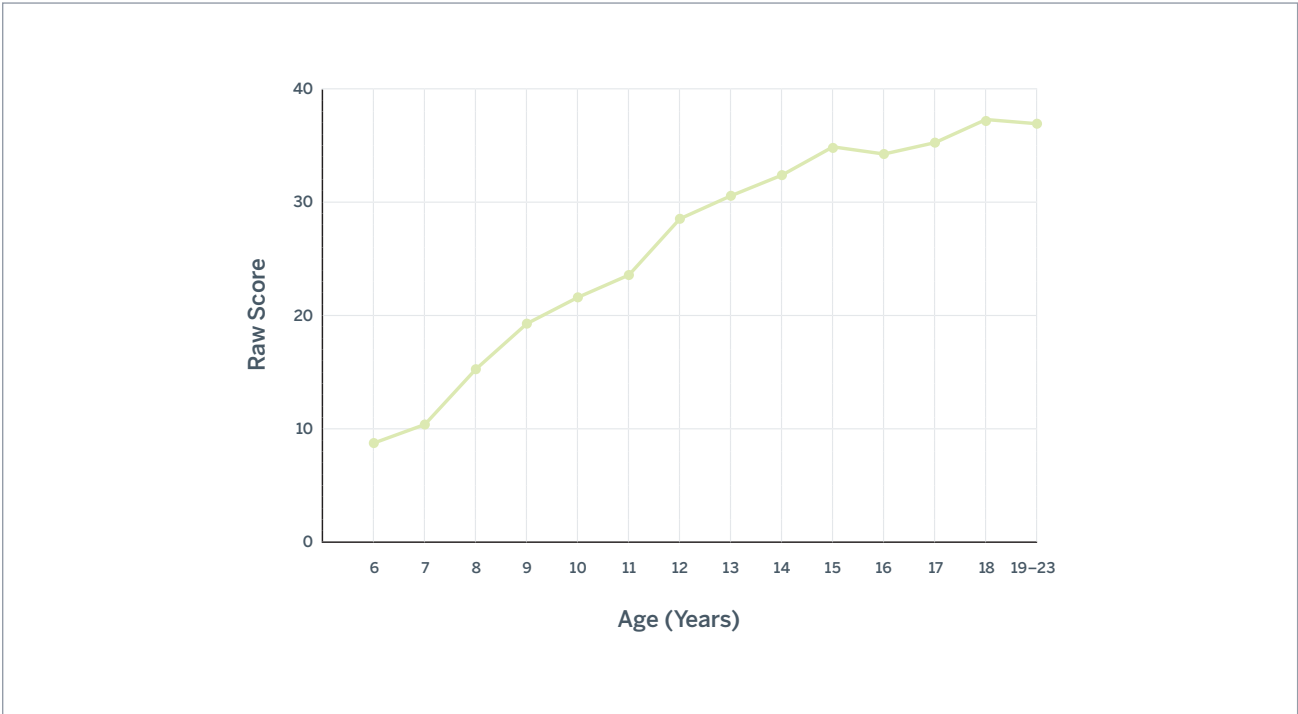
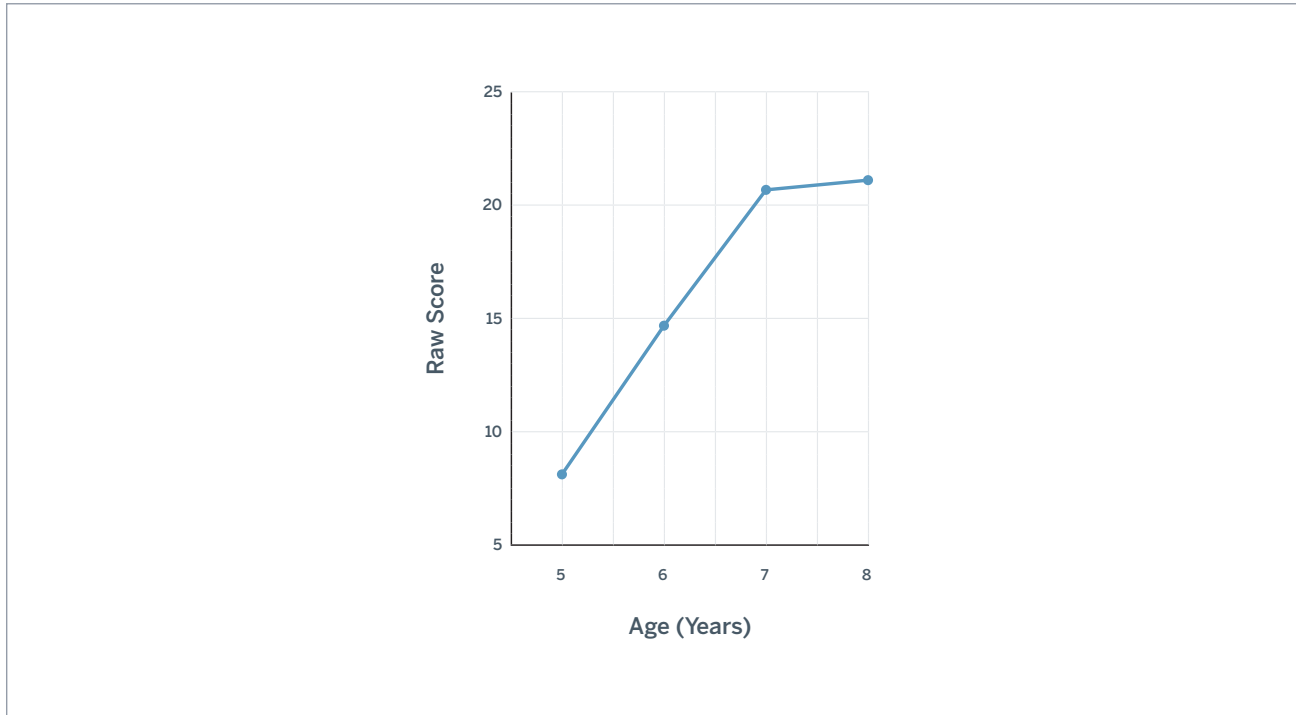


Figure 5.4. Growth Beyond High School Example: Regular Word Spelling (Test 15C)

### TOD-E

All six TOD-E tests demonstrate a similar progression of means, as illustrated by the Rhyming (5E) test examples shown in Figure 5.5. The TOD-E was designed to measure earlier skills displayed in kindergarten and first grade by beginning readers; therefore, a steeper initial progression of scores

followed by a leveling off toward the top of the age range fits the design and expectations of the tests. Since the TOD-E is meant to identify struggling readers, it was important to include second graders even though typically developing second graders reached a ceiling on the TOD-E tests.



*Figure 5.5. TOD-E Ceilings Example: Rhyming (Test 5E)*

### Test Intercorrelations

The correlations between the individual TOD test scores in the standardization sample were examined to provide further evidence of construct validity. Tables 5.14–5.18 display the intercorrelations for tests in the standardization samples: TOD-S child, TOD-S adult, TOD-C child, TOD-C adult, and TOD-E. As expected, the tests exhibit correlations that range considerably, from small (.08) to high (.84). Lower correlations were found between tests of divergent skill areas. For example, Phonological Manipulation (4C) and Rapid Number and Letter Naming (17C) correlate at .31 in the TOD-C child sample and .30 in the TOD-C adult sample; phonological awareness and rapid automatized naming would not be expected to correlate highly. Higher correlations were found between tests of similar skills, such as Regular Word Spelling (15C) and Irregular Word Spelling (5C), which correlate at .81 in the TOD-C child sample and .78 in the TOD-C adult sample.

For all tests, the correlations are lower than their internal consistency reliabilities reported earlier in the chapter. The correlations are, however, high enough to warrant their combination to produce index and composite scores (of combined test scores) and low enough to show that each test measures a unique skill and thus can be scored and interpreted independently.

**Table 5.14.** Test Intercorrelations: TOD-S Child

	PV-S	LWC-S	WRF-S	QRF-S
PV-S	—			
LWC-S	.57	—		
WRF-S	.39	.55	—	
QRF-S	.42	.66	NA	—

Note.  $N = 1,723$ . PV-S = Picture Vocabulary; LWC-S = Letter and Word Choice; WRF-S = Word Reading Fluency; QRF-S = Question Reading Fluency.

**Table 5.15.** Test Intercorrelations: TOD-S Adult

	PV-S	LWC-S	QRF-S
PV-S	—		
LWC-S	.58	—	
QRF-S	.39	.40	—

Note.  $N = 347$ . PV-S = Picture Vocabulary; LWC-S = Letter and Word Choice; QRF-S = Question Reading Fluency.

**Table 5.16.** Test Intercorrelations: TOD-C Child (Including TOD-S)

	PV-S	LWC-S	WRF-S	QRF-S	PHM-C	IWS-C	RLN-C	PWR-C	WPC-C	WM-C	PAN-C	IWR-C
PV-S	—											
LWC-S	.57	—										
WRF-S	.53	.72	—									
QRF-S	.39	.53	NA	—								
PHM-C	.28	.33	.59	.36	—							
IWS-C	.46	.68	.65	.54	.44	—						
RLN-C	.27	.41	.72	.47	.30	.49	—					
PWR-C	.36	.48	.74	.48	.47	.62	.52	—				
WPC-C	.22	.33	.47	.43	.23	.41	.36	.32	—			
WM-C	.32	.30	.44	.29	.34	.39	.28	.33	.20	—		
PAN-C	.34	.27	.08	.31	.30	.26	.23	.34	.17	.29	—	
IWR-C	.49	.63	.81	.53	.39	.68	.49	.61	.31	.31	.32	—
ORE-C	.29	.43	.82	.51	.59	.52	.44	.49	.29	.24	.22	.52
BLN-C	.29	.25	.28	.24	.34	.34	.25	.33	.16	.28	.30	.33
SEG-C	.25	.20	.35	.19	.35	.25	.18	.31	.11	.29	.33	.31
RWS-C	.46	.64	.84	.54	.48	.81	.51	.68	.38	.43	.33	.68
SRE1-C	.45	.57	.75	.65	.46	.60	.45	.53	.44	.31	.27	.63
SRE2-C	.50	.57	NA	.69	.47	.60	.44	.51	.44	.35	.38	.55
RNL-C	.22	.32	.65	.45	.31	.38	.66	.53	.32	.23	.25	.42
LM-C	.26	.32	.42	.29	.30	.43	.27	.33	.24	.62	.23	.33
RPW-C	.36	.52	.78	.54	.46	.65	.59	.84	.35	.38	.30	.65
RIW-C	.36	.52	.84	.60	.37	.59	.60	.67	.37	.25	.30	.66
SSL-C	.27	.25	.33	.26	.34	.30	.24	.40	.18	.37	.27	.29
LV-C	.56	.49	.42	.50	.43	.52	.34	.49	.27	.38	.45	.56
GAN-C	.41	.36	.53	.36	.38	.39	.27	.43	.16	.42	.49	.37

Note.  $N = 1,401$ . PV-S = Picture Vocabulary; LWC-S = Letter and Word Choice; WRF-S = Word Reading Fluency; QRF-S = Question Reading Fluency; PHM-C = Phonological Manipulation; IWS-C = Irregular Word Spelling; RLN-C = Rapid Letter Naming; PWR-C = Pseudoword Reading; WPC-C = Word Pattern Choice; WM-C = Word Memory; PAN-C = Picture Analogies; IWR-C = Irregular Word Reading; ORE-C = Oral Reading Efficiency; BLN-C = Blending; SEG-C = Segmenting; RWS-C = Regular Word Spelling; SRE1-C = Silent Reading Efficiency Grades 1–5; SRE2-C = Silent Reading Efficiency Grade 6–Adult; RNL-C = Rapid Number and Letter Naming; LM-C = Letter Memory; RPW-C = Rapid Pseudoword Reading; RIW-C = Rapid Irregular Word Reading; SSL-C = Symbol to Sound Learning; LV-C = Listening Vocabulary; GAN-C = Geometric Analogies.



**Table 5.16.** Test Intercorrelations: TOD-C Child (Including TOD-S) (*continued*)

	ORE-C	BLN-C	SEG-C	RWS-C	SRE1-C	SRE2-C	RNL-C	LM-C	RPW-C	RIW-C	SSL-C	LV-C	GAN-C
PV-S													
LWC-S													
WRF-S													
QRF-S													
PHM-C													
IWS-C													
RLN-C													
PWR-C													
WPC-C													
WM-C													
PAN-C													
IWR-C													
ORE-C	—												
BLN-C	.21	—											
SEG-C	.19	.58	—										
RWS-C	.50	.39	.34	—									
SRE1-C	.65	.22	.20	.59	—								
SRE2-C	.51	.28	.21	.63	NA	—							
RNL-C	.42	.22	.20	.44	.45	.38	—						
LM-C	.26	.26	.27	.44	.32	.34	.23	—					
RPW-C	.55	.31	.28	.71	.58	.54	.59	.37	—				
RIW-C	.55	.24	.20	.60	.63	.58	.62	.24	.72	—			
SSL-C	.20	.20	.25	.35	.29	.28	.27	.36	.37	.31	—		
LV-C	.38	.38	.38	.56	.49	.64	.32	.34	.49	.48	.33	—	
GAN-C	.27	.34	.35	.46	.36	.47	.28	.36	.40	.35	.36	.49	—

Note. *N* = 1,401. **PV-S** = Picture Vocabulary; **LWC-S** = Letter and Word Choice; **WRF-S** = Word Reading Fluency; **QRF-S** = Question Reading Fluency; **PHM-C** = Phonological Manipulation; **IWS-C** = Irregular Word Spelling; **RLN-C** = Rapid Letter Naming; **PWR-C** = Pseudoword Reading; **WPC-C** = Word Pattern Choice; **WM-C** = Word Memory; **PAN-C** = Picture Analogies; **IWR-C** = Irregular Word Reading; **ORE-C** = Oral Reading Efficiency; **BLN-C** = Blending; **SEG-C** = Segmenting; **RWS-C** = Regular Word Spelling; **SRE1-C** = Silent Reading Efficiency Grades 1–5; **SRE2-C** = Silent Reading Efficiency Grade 6–Adult; **RNL-C** = Rapid Number and Letter Naming; **LM-C** = Letter Memory; **RPW-C** = Rapid Pseudoword Reading; **RIW-C** = Rapid Irregular Word Reading; **SSL-C** = Symbol to Sound Learning; **LV-C** = Listening Vocabulary; **GAN-C** = Geometric Analogies.

**Table 5.17.** Test Intercorrelations: TOD-C Adult (Including TOD-S)

	PV-S	LWC-S	QRF-S	PHM-C	IWS-C	RLN-C	PWR-C	WPC-C	WM-C	PAN-C	IWR-C
PV-S	—										
LWC-S	.58	—									
QRF-S	.39	.40	—								
PHM-C	.55	.47	.38	—							
IWS-C	.56	.66	.53	.50	—						
RLN-C	.32	.38	.47	.36	.46	—					
PWR-C	.51	.49	.42	.69	.63	.54	—				
WPC-C	.33	.39	.43	.40	.36	.29	.38	—			
WM-C	.49	.43	.34	.49	.49	.40	.45	.32	—		
PAN-C	.42	.32	.40	.44	.37	.35	.38	.30	.41	—	
IWR-C	.60	.58	.45	.49	.73	.50	.70	.36	.49	.44	—
ORE-C	.38	.42	.51	.43	.56	.50	.43	.32	.36	.33	.52
BLN-C	.38	.34	.33	.58	.32	.21	.37	.26	.43	.38	.28
SEG-C	.44	.43	.31	.63	.42	.32	.53	.34	.48	.41	.47
RWS-C	.60	.63	.52	.60	.78	.52	.71	.44	.53	.43	.72
SRE2-C	.58	.50	.69	.51	.60	.48	.56	.47	.47	.45	.60
RNL-C	.30	.35	.47	.30	.47	.76	.50	.31	.33	.32	.49
LM-C	.40	.41	.38	.46	.46	.37	.47	.29	.65	.41	.48
RPW-C	.49	.58	.43	.56	.65	.57	.82	.40	.45	.32	.70
RIW-C	.45	.44	.52	.45	.58	.62	.61	.40	.37	.39	.59
SSL-C	.50	.45	.31	.54	.42	.29	.47	.27	.53	.41	.47
LV-C	.65	.51	.53	.58	.58	.44	.59	.40	.51	.52	.66
GAN-C	.53	.37	.36	.55	.46	.40	.51	.32	.53	.66	.51

Note.  $N = 347$ . WRF-S and SRE1-C are not taken by adults. PV-S = Picture Vocabulary; LWC-S = Letter and Word Choice; QRF-S = Question Reading Fluency; PHM-C = Phonological Manipulation; IWS-C = Irregular Word Spelling; RLN-C = Rapid Letter Naming; PWR-C = Pseudoword Reading; WPC-C = Word Pattern Choice; WM-C = Word Memory; PAN-C = Picture Analogies; IWR-C = Irregular Word Reading; ORE-C = Oral Reading Efficiency; BLN-C = Blending; SEG-C = Segmenting; RWS-C = Regular Word Spelling; SRE2-C = Silent Reading Efficiency Grade 6–Adult; RNL-C = Rapid Number and Letter Naming; LM-C = Letter Memory; RPW-C = Rapid Pseudoword Reading; RIW-C = Rapid Irregular Word Reading; SSL-C = Symbol to Sound Learning; LV-C = Listening Vocabulary; GAN-C = Geometric Analogies; WRF-S = Word Reading Fluency; SRE1-C = Silent Reading Efficiency Grades 1–5.

**Table 5.17.** Test Intercorrelations: TOD-C Adult (Including TOD-S) (*continued*)

	ORE-C	BLN-C	SEG-C	RWS-C	SRE2-C	RNL-C	LM-C	RPW-C	RIW-C	SSL-C	LV-C	GAN-C
PV-S												
LWC-S												
QRF-S												
PHM-C												
IWS-C												
RLN-C												
PWR-C												
WPC-C												
WM-C												
PAN-C												
IWR-C												
ORE-C	—											
BLN-C	.29	—										
SEG-C	.33	.64	—									
RWS-C	.54	.38	.49	—								
SRE2-C	.58	.39	.44	.62	—							
RNL-C	.50	.17	.25	.53	.50	—						
LM-C	.39	.43	.47	.51	.40	.30	—					
RPW-C	.50	.31	.47	.71	.55	.56	.44	—				
RIW-C	.45	.28	.37	.62	.61	.68	.37	.63	—			
SSL-C	.24	.41	.46	.48	.36	.19	.47	.42	.28	—		
LV-C	.56	.44	.55	.67	.66	.42	.52	.59	.53	.48	—	
GAN-C	.32	.42	.51	.49	.44	.30	.45	.42	.43	.53	.56	—

Note. *N* = 347. WRF-S and SRE1-C are not taken by adults. **PV-S** = Picture Vocabulary; **LWC-S** = Letter and Word Choice; **QRF-S** = Question Reading Fluency; **PHM-C** = Phonological Manipulation; **IWS-C** = Irregular Word Spelling; **RLN-C** = Rapid Letter Naming; **PWR-C** = Pseudoword Reading; **WPC-C** = Word Pattern Choice; **WM-C** = Word Memory; **PAN-C** = Picture Analogies; **IWR-C** = Irregular Word Reading; **ORE-C** = Oral Reading Efficiency; **BLN-C** = Blending; **SEG-C** = Segmenting; **RWS-C** = Regular Word Spelling; **SRE2-C** = Silent Reading Efficiency Grade 6–Adult; **RNL-C** = Rapid Number and Letter Naming; **LM-C** = Letter Memory; **RPW-C** = Rapid Pseudoword Reading; **RIW-C** = Rapid Irregular Word Reading; **SSL-C** = Symbol to Sound Learning; **LV-C** = Listening Vocabulary; **GAN-C** = Geometric Analogies; **WRF-S** = Word Reading Fluency; **SRE1-C** = Silent Reading Efficiency Grades 1–5.

**Table 5.18.** Test Intercorrelations: TOD-E (Including TOD-S)

	PV-S	LWC-S	WRF-S	QRF-S	SPW-E	RHY-E	ERNL-E	LSW-E	ESEG-E	LSK-E
PV-S	—									
LWC-S	.58	—								
WRF-S	.37	.64	—							
QRF-S	.42	.72	NA	—						
SPW-E	.47	.65	.61	.57	—					
RHY-E	.42	.52	.39	.49	.61	—				
ERNL-E	.37	.40	.42	.45	.56	.46	—			
LSW-E	.47	.66	.68	.70	.76	.57	.64	—		
ESEG-E	.32	.22	.18	.10	.45	.39	.46	.38	—	
LSK-E	.45	.55	.51	.62	.72	.60	.59	.73	.52	—

Note. *N* = 342. PV-S = Picture Vocabulary; LWC-S = Letter and Word Choice; WRF-S = Word Reading Fluency; QRF-S = Question Reading Fluency; SPW-E = Sounds and Pseudowords; RHY-E = Rhyming; ERNL-E = Early Rapid Number and Letter Naming; LSW-E = Letter and Sight Word Recognition; ESEG-E = Early Segmenting; LSK-E = Letter and Sound Knowledge.

### Confirmatory Factor Analysis (CFA) Evidence Supporting the TOD Diagnostic Indexes

As described in earlier sections, the TOD tests were designed to measure the hallmark linguistic risk factors of dyslexia, specifically limited phonological awareness, poor orthographic processing, slow rapid automatized naming, and limited working memory, along with reading and spelling skills that are typically negatively impacted by dyslexia (Mather & Wendling, in press; McCallum et al., 2006). The Dyslexia Diagnostic Index (DDI) and Early Dyslexia Diagnostic Index (EDDI) were created based on this research as well as a series of multiple regression analyses, described in Chapter 4. Because these diagnostic indicators are the most powerful (global) predictors within the TOD-C and TOD-E, they were subjected to Confirmatory Factor Analysis (CFA) model testing. That is, two theoretical models were hypothesized to explain these scores in order to evaluate the utility of the overall diagnostic indexes: a one-factor model whereby all tests load onto the overall diagnostic index; and a two-factor model whereby the tests are separated into the groups that make up two component scores, the Linguistic Processing Index (LPI) for the TOD-C and Early Linguistic Processing Index (ELPI) for the TOD-E; and

the Reading and Spelling Index (RSI) for the TOD-C and Early Reading and Spelling Index (ERSI) for the TOD-E. CFA was applied to evaluate the models, with modification if necessary, and analyzed using Mplus (Version 7) software (Muthén & Muthén, 2012). This analytic approach compares the goodness-of-fit statistics of the one- and two-factor models to evaluate the extent to which these hypothesized models fit the sample data (Byrnes, 2012).

### Standardization Sample CFA

Data defining these factor structures of the TOD-C and TOD-E were taken from the standardization samples described in Chapter 4. Tables 5.19 and 5.20 present the model fit statistics, along with factor loadings and factor correlations, for the TOD-C and TOD-E samples. For both samples, the goodness-of-fit statistics represent an acceptable fit of the models across the standardization samples. Further, the one-factor and two-factor models are virtually the same, i.e., the two-factor model does not significantly improve the model fit compared to the one-factor model, and thus interpretation is appropriate using either the one-factor or two-factor model.

**Table 5.19.** Comparing Confirmatory Factor Analysis Model Fit for the TOD-C Standardization Sample

	One-factor model	Two-factor model	
<b>Model fit statistics</b>			
chi-square	456	452	
<i>df</i>	20	19	
<i>p</i>	<.001	<.001	
SRMR	.04	.04	
RMSEA	.11	.11	
CFI	.92	.92	
TLI	.89	.88	
<hr/>			
<b>Factor loadings</b>	<b>Dyslexia Diagnostic Index (DDI)</b>	<b>Linguistic Processing Index (LPI)</b>	<b>Reading and Spelling Index (RSI)</b>
Letter and Word Choice	.72	—	.72
Word/Question Reading Fluency	.68	—	.68
Phonological Manipulation	.70	.71	—
Irregular Word Spelling	.83	—	.84
Rapid Letter Naming	.62	.62	—
Pseudoword Reading	.75	—	.75
Word Pattern Choice	.49	.49	—
Word Memory	.51	.52	—

Note. *N* = 1,748. *df* = degrees of freedom; *p* = the probability, testing against the null hypothesis, that the RMSEA is zero; **SRMR** = standardized root-mean-square residual, average correlation residuals; **RMSEA** = root-mean-square error of approximation, function of chi-square test of close fit; **CFI** = comparative fit index; **TLI** = Tucker–Lewis index.

**Table 5.20.** Comparing Confirmatory Factor Analysis Model Fit for the TOD-E Standardization Sample

	One-factor model	Two-factor model	
<b>Model fit statistics</b>			
chi-square	149	143	
<i>df</i>	20	19	
<i>p</i>	<.001	<.001	
SRMR	.05	.05	
RMSEA	.13	.13	
CFI	.92	.92	
TLI	.89	.89	
<hr/>			
<b>Factor loadings</b>	<b>Early Dyslexia Diagnostic Index (EDDI)</b>	<b>Early Linguistic Processing Index (ELPI)</b>	<b>Early Reading and Spelling Index (ERSI)</b>
Letter and Word Choice	.73	—	.73
Word/Question Reading Fluency	.70	—	.70
Sounds and Pseudowords	.86	—	.86
Rhyming	.68	.71	—
Early Rapid Number and Letter Naming	.68	.72	—
Letter and Sight Word Recognition	.89	—	.89
Early Segmenting	.49	.53	—
Letter and Sound Knowledge	.83	—	.82

Note. *N* = 342. *df* = degrees of freedom; *p* = the probability, testing against the null hypothesis, that the RMSEA is zero; **SRMR** = standardized root-mean-square residual, average correlation residuals; **RMSEA** = root-mean-square error of approximation, function of chi-square test of close fit; **CFI** = comparative fit index; **TLI** = Tucker–Lewis index.

### Clinical Sample CFA

For cross-validation purposes, the same theoretical models were examined in the clinical samples for the TOD-C and TOD-E. Due to the lower prevalence of identified disorders in the adult sample, only the child data were included in the clinical data set for the TOD-C. Tables 5.21 and 5.22 present the model fit statistics, along with factor loadings and factor

correlations, for the TOD-C and TOD-E clinical samples. Again, for both samples, the goodness-of-fit statistics represent an acceptable fit of the models, and the two-factor model does not significantly improve the model fit compared to the one-factor model. Additionally, the model fit statistics and factor loadings are slightly strengthened in the clinical sample when compared to the standardization sample.

**Table 5.21.** Comparing Confirmatory Factor Analysis Model Fit for the TOD-C Clinical Sample

	One-factor model	Two-factor model	
<b>Model fit statistics</b>			
chi-square	229		227
<i>df</i>	20		19
<i>p</i>	<.001		<.001
SRMR	.04		.04
RMSEA	.14		.15
CFI	.91		.91
TLI	.87		.87
<hr/>			
<b>Factor loadings</b>	<b>Dyslexia Diagnostic Index (DDI)</b>	<b>Linguistic Processing Index (LPI)</b>	<b>Reading and Spelling Index (RSI)</b>
Letter and Word Choice	.83	—	.83
Word/Question Reading Fluency	.83	—	.83
Phonological Manipulation	.74	.75	—
Irregular Word Spelling	.80	—	.80
Rapid Letter Naming	.79	.80	—
Pseudoword Reading	.79	—	.79
Word Pattern Choice	.61	.62	—
Word Memory	.45	.46	—

Note. *N* = 515. *df* = degrees of freedom; *p* = the probability, testing against the null hypothesis, that the RMSEA is zero; SRMR = standardized root-mean-square residual, average correlation residuals; RMSEA = root-mean-square error of approximation, function of chi-square test of close fit; CFI = comparative fit index; TLI = Tucker–Lewis index.

**Table 5.22.** Comparing Confirmatory Factor Analysis Model Fit for the TOD-E Clinical Sample

	One-factor model	Two-factor model	
<b>Model fit statistics</b>			
chi-square	34	34	
<i>df</i>	20	19	
<i>p</i>	<.028	<.021	
SRMR	.04	.04	
RMSEA	.10	.11	
CFI	.97	.96	
TLI	.95	.95	
<hr/>			
<b>Factor loadings</b>	<b>Early Dyslexia Diagnostic Index (EDDI)</b>	<b>Early Linguistic Processing Index (ELPI)</b>	<b>Early Reading and Spelling Index (ERSI)</b>
Letter and Word Choice	.82	—	.82
Word/Question Reading Fluency	.68	—	.68
Sounds and Pseudowords	.90	—	.89
Rhyming	.70	.69	—
Early Rapid Number and Letter Naming	.79	.78	—
Letter and Sight Word Recognition	.93	—	.93
Early Segmenting	.66	.65	—
Letter and Sound Knowledge	.86	—	.87

Note. *N* = 68. *df* = degrees of freedom; *p* = the probability, testing against the null hypothesis, that the RMSEA is zero; **SRMR** = standardized root-mean-square residual, average correlation residuals; **RMSEA** = root-mean-square error of approximation, function of chi-square test of close fit; **CFI** = comparative fit index; **TLI** = Tucker–Lewis index.



## Convergent Validity

The *convergent validation* method examines a test's relationship to other measures of similar constructs. It is sometimes referred to as *concurrent validity*. Moderate to strong correlations with convergent measures are seen as supporting the construct validity of the test under study. This section describes the related assessments that were administered for the convergent validity study and their correlations with the TOD tests that assess similar constructs/skills. The TOD-S tests are not presented separately but

are included with the TOD-C and TOD-E analyses. Due to the large number of TOD tests, only the most relevant correlations are presented. These analyses focused on comparing tests of similar skills/constructs to one another and therefore do not involve the index or composite scores.

Each related assessment was taken by a subset of individuals from the standardization and clinical samples. Tables 5.23 (TOD-S/TOD-C) and, later in this chapter, 5.25 (TOD-S/TOD-E) present the demographic characteristics for these validation groups.

**Table 5.23.** Demographic Characteristics of the TOD-S/TOD-C Convergent Validation Samples

	WJ IV COG	WJ IV ACH	CASL-2	CTOPP-2	TOWRE-2	TOC-2	UNIT-GAT
<b>Total sample</b>	48	57	49	48	46	46	42
<b>Gender</b>							
Male	22	28	28	16	26	21	14
Female	26	29	21	32	20	25	28
<b>Parents' /Individual's educational level</b>							
No high school diploma	2	3	0	1	0	6	1
High school graduate	12	9	3	8	15	10	15
Some college	13	15	14	10	7	13	3
Bachelor's degree or higher	21	30	32	29	24	17	23
<b>Race/Ethnicity<sup>a</sup></b>							
Asian	0	4	2	1	5	2	8
Black/African American	0	2	1	1	2	2	3
White	17	39	34	23	25	21	12
American Indian/Alaska Native	0	0	1	0	7	1	1
Native Hawaiian/Pacific Islander	1	1	0	0	2	0	0
Other/Multiracial	0	0	4	3	0	1	0
Hispanic Origin	30	11	7	20	5	19	18
<b>U.S. geographic region</b>							
Northeast	0	5	4	0	0	15	0
Midwest	8	10	30	26	18	6	0
South	31	32	15	14	28	25	42
West	9	10	0	8	0	0	0

Note. **WJ IV COG** = Woodcock-Johnson IV Tests of Cognitive Abilities; **WJ IV ACH** = Woodcock-Johnson IV Tests of Achievement; **CASL-2** = Comprehensive Assessment of Spoken Language, Second Edition; **CTOPP-2** = Comprehensive Test of Phonological Processing, Second Edition; **TOWRE-2** = Test of Word Reading Efficiency, Second Edition; **TOC-2** = Test of Orthographic Competence, Second Edition; **UNIT-GAT** = Universal Nonverbal Intelligence Test–Group Abilities Test.

<sup>a</sup>Individuals of Hispanic origin are included in the race/ethnicity category under Hispanic Origin; remaining categories include only individuals of non-Hispanic origin.

**Table 5.23.** Demographic Characteristics of the TOD-S/TOD-C Convergent Validation Samples (*continued*)

	WJ IV COG	WJ IV ACH	CASL-2	CTOPP-2	TOWRE-2	TOC-2	UNIT-GAT
<b>Age (years)</b>							
6	0	0	0	0	1	0	0
7	1	1	0	0	0	0	0
8	3	3	3	2	5	3	5
9	2	4	5	8	1	3	1
10	6	8	10	10	8	5	3
11	3	3	5	2	6	3	2
12	5	10	6	3	5	7	2
13	4	4	4	4	3	4	2
14	2	5	3	2	2	10	3
15	6	5	4	7	1	4	4
16	3	2	3	4	3	1	4
17	2	2	2	6	1	2	4
18	0	0	0	0	3	0	1
19–23	2	2	4	0	7	4	11
24–89	9	8	0	0	0	0	0
<b>Disability status</b>							
Clinical	10	23	26	15	12	9	10
Typical	38	34	23	33	34	37	32

Note. **WJ IV COG** = Woodcock-Johnson IV Tests of Cognitive Abilities; **WJ IV ACH** = Woodcock-Johnson IV Tests of Achievement; **CASL-2** = Comprehensive Assessment of Spoken Language, Second Edition; **CTOPP-2** = Comprehensive Test of Phonological Processing, Second Edition; **TOWRE-2** = Test of Word Reading Efficiency, Second Edition; **TOC-2** = Test of Orthographic Competence, Second Edition; **UNIT-GAT** = Universal Nonverbal Intelligence Test–Group Abilities Test.

<sup>a</sup>Individuals of Hispanic origin are included in the race/ethnicity category under Hispanic Origin; remaining categories include only individuals of non-Hispanic origin.

## TOD-C

For the TOD-C, convergent validity data were collected from seven assessments: Comprehensive Assessment of Spoken Language, Second Edition (CASL-2; Carrow-Woolfolk, 2017); Test of Orthographic Competence, Second Edition (TOC-2; Mather et al., 2022); Woodcock-Johnson IV Tests of Achievement (WJ IV ACH; Schrank et al., 2014a); Comprehensive Test of Phonological Processing, Second Edition (CTOPP-2; Wagner et al., 2013); Woodcock-Johnson IV Tests of Cognitive Abilities (WJ IV COG; Schrank et al., 2014b); Universal Nonverbal Intelligence Test–Group Abilities Test (UNIT-GAT; Bracken & McCallum, 2019); Test of Word Reading Efficiency, Second Edition (TOWRE-2; Torgesen et al., 2012). Table 5.24 presents the correlations between the TOD-S and TOD-C tests and their corresponding validation tests. Correlations ranged from .30 to .91, and most are moderate. As a reminder, a correlation of 0.2 is considered small, 0.5 is considered medium/moderate, and 0.8 is considered large/strong (Cohen, 1992).

## CASL-2

The CASL-2 is an individually administered test of spoken language. The Receptive Vocabulary test of the CASL-2 was administered to a sample of 49 individuals from the TOD-C standardization and clinical samples. The Receptive Vocabulary test, which requires the examinee to choose which image matches the word that the examiner says aloud, was used to provide validation support for the TOD-S Picture Vocabulary (1S) and the TOD-C Listening Vocabulary (22C) tests. Both TOD tests demonstrated a moderate relationship with CASL-2 Receptive Vocabulary.

## TOC-2

The TOC-2 assesses orthographic processing skills that are integral to proficient reading and writing. Two subtests of the TOC-2, Homophone Spelling and Letter Choice, were taken by 46 individuals from the TOD-C standardization and clinical samples. TOC-2 Homophone Spelling, which requires the examinee to provide the correct spelling of a homophone presented by a picture, was used to validate TOD-S Letter and Word Choice (2S). TOC-2 Letter Choice is a timed test in which the examinee completes words with a missing letter (*b*, *d*, *p*, or *q*) and was compared with TOD-S Letter and Word Choice (2S), as well as TOD-C Word Pattern Choice (8C). All correlations

were moderate to strong, providing validation support for the TOD by the TOC-2 orthographic processing tests.

## WJ IV Tests of Achievement

The WJ IV ACH measures academic achievement skills. For the purposes of this study, 57 individuals from the TOD-C standardization and clinical samples took five WJ IV ACH tests: Spelling, Letter–Word Identification, Word Attack, Passage Comprehension, and Sentence Reading Fluency. The WJ IV ACH Spelling test correlated highly with both TOD-C tests of spelling (Irregular Word Spelling [5C], Regular Word Spelling [15C]) as well as TOD-S Letter and Word Choice (2S), a test of spelling recognition. The Letter–Word Identification and Word Attack tests correlated at a moderate to high level with the TOD-C tests of word reading (Pseudoword Reading [7C], Irregular Word Reading [11C]), and the Letter–Word Identification test also correlated highly with Letter and Word Choice (2S). The WJ IV ACH Passage Comprehension test correlated at a moderate level with both TOD-C reading comprehension tests (Question Reading Fluency [3Sb], Silent Reading Efficiency [16C]). The WJ IV ACH Sentence Reading Fluency test also correlated at a moderate level with these two tests, as well as with TOD-C Oral Reading Efficiency (12C). These moderate to high correlations between TOD tests and similar constructs on the WJ IV ACH provide validation support for the TOD.

## CTOPP-2

The CTOPP-2 is an individually administered test of phonological skills used to determine whether an individual is at risk for reading difficulties. Forty-eight individuals from the TOD-C standardization and clinical samples were administered five subtests from the CTOPP-2: Elision, Blending Words, Phoneme Isolation, Rapid Digit Naming, and Rapid Letter Naming. The three CTOPP-2 tests of phonological awareness (Elision, Blending Words, and Phoneme Isolation) correlated at a low to moderate level with the TOD-C tests of phonological awareness (Phonological Manipulation [4C], Blending [13C], and Segmenting [14C]). The two CTOPP-2 rapid naming tests correlated moderately with the two TOD-C rapid naming tests (Rapid Letter Naming [6C], Rapid Number and Letter Naming [17C]).

Table 5.24. Convergent Validation Correlations: TOD-S/TOD-C

TOD-S/TOD-C Test <sup>a</sup>	CASL-2			TOC-2			WJ IV ACH		
	Receptive Vocabulary	Homophone Spelling	Letter Choice	Spelling	Letter-Word ID	Word Attack	Passage Comprehension	Sentence Reading Fluency	
Picture Vocabulary (1S)	.49								
Letter and Word Choice (2S)		.74	.56	.81	.65		.43	.54	
Question Reading Fluency (3Sb)									
Phonological Manipulation (4C)				.90					
Irregular Word Spelling (5C)					.65	.73			
Rapid Letter Naming (6C)									
Pseudoword Reading (7C)									
Word Pattern Choice (8C)			.46						
Word Memory (9C)									
Picture Analogies (10C)									
Irregular Word Reading (11C)					.76	.72		.45	
Oral Reading Efficiency (12C)									
Blending (13C)									
Segmenting (14C)									
Regular Word Spelling (15C)				.91					
Silent Reading Efficiency Grade 6-Adult (16C)							.52	.69	
Rapid Number and Letter Naming (17C)									
Letter Memory (18C)									
Rapid Pseudoword Reading (19C)									
Rapid Irregular Word Reading (20C)									
Symbol to Sound Learning (21C)									
Listening Vocabulary (22C)	.51								
Geometric Analogies (23C)									

Note. *n* varies by test: CASL-2 *n* = 49; TOC-2 *n* = 46; WJ IV ACH *n* = 57; CTOPP-2 *n* = 48; WJ IV COG *n* = 48; UNIT-GAT *n* = 42; TOWRE-2 *n* = 46. All correlations significant at  $p < .01$  except CTOPP-2 Elision, Blending Words, and Phoneme Isolation with TOD-C Blending and Segmenting significant at  $p > .05$ . CASL-2 = Comprehensive Assessment of Spoken Language, Second Edition; TOC-2 = Test of Orthographic Competence, Second Edition; WJ IV ACH = Woodcock-Johnson IV Tests of Achievement; CTOPP-2 = Comprehensive Test of Phonological Processing, Second Edition; WJ IV COG = Woodcock-Johnson IV Tests of Cognitive Abilities; UNIT-GAT = Universal Nonverbal Intelligence Test-Group Abilities Test; TOWRE-2 = Test of Word Reading Efficiency, Second Edition.

<sup>a</sup>Sample sizes for Word Reading Fluency and Silent Reading Efficiency Grades 1-5 were too small for correlational analyses.

Table 5.24 continued on next page

Table 5.24. Convergent Validation Correlations: TOD-S/TOD-C (continued)

TOD-S/TOD-C Test <sup>a</sup>	CTOPP-2					WJ IV COG			UNIT-GAT		TOWRE-2
	Elision	Blending Words	Phoneme Isolation	Rapid Digit Naming	Rapid Letter Naming	Numbers Reversed	Visual Auditory Learning	Object Number Sequencing	Full Scale	Full Scale	Full Scale
Picture Vocabulary (1S)											
Letter and Word Choice (2S)											
Question Reading Fluency (3Sb)											
Phonological Manipulation (4C)	.61	.45	.44								
Irregular Word Spelling (5C)											
Rapid Letter Naming (6C)				.54	.42						
Pseudoword Reading (7C)											
Word Pattern Choice (8C)											
Word Memory (9C)						.57		.53			
Picture Analogies (10C)									.47		
Irregular Word Reading (11C)											
Oral Reading Efficiency (12C)											.55
Blending (13C)	.30	.50	.38								
Segmenting (14C)	.36	.35	.34								
Regular Word Spelling (15C)											
Silent Reading Efficiency Grade 6-Adult (16C)											
Rapid Number and Letter Naming (17C)				.56	.59						
Letter Memory (18C)						.57		.46			
Rapid Pseudoword Reading (19C)											.82
Rapid Irregular Word Reading (20C)											.82
Symbol to Sound Learning (21C)							.56				
Listening Vocabulary (22C)											
Geometric Analogies (23C)										.58	

Note. *n* varies by test: CASL-2 *n* = 49; TOC-2 *n* = 46; WJ IV ACH *n* = 57; CTOPP-2 *n* = 48; WJ IV COG *n* = 48; UNIT-GAT *n* = 42; TOWRE-2 *n* = 46. All correlations significant at <.01 except CTOPP-2 Elision, Blending Words, and Phoneme Isolation with TOD-C Blending and Segmenting significant at >.05. CASL-2 = Comprehensive Assessment of Spoken Language, Second Edition; TOC-2 = Test of Orthographic Competence, Second Edition; WJ IV ACH = Woodcock-Johnson IV Tests of Achievement; CTOPP-2 = Comprehensive Test of Phonological Processing, Second Edition; WJ IV COG = Woodcock-Johnson IV Tests of Cognitive Abilities; UNIT-GAT = Universal Nonverbal Intelligence Test-Group Abilities Test; TOWRE-2 = Test of Word Reading Efficiency, Second Edition.

<sup>a</sup>Sample sizes for Word Reading Fluency and Silent Reading Efficiency Grades 1-5 were too small for correlational analyses.

### **WJ IV Tests of Cognitive Abilities**

Three subtests from the WJ IV COG were taken by 48 individuals from the TOD-C standardization and clinical samples. Two of these tests, Numbers Reversed and Object Number Sequencing, measure auditory working memory and demonstrated moderate correlations with the two TOD-C tests of working memory, Word Memory (9C) and Letter Memory (18C). The Visual Auditory Learning test on the WJ IV COG correlated moderately with Symbol to Sound Learning (21C), both of which are tests of visual-verbal paired-associate learning.

### **UNIT-GAT**

The UNIT-GAT is a nonverbal screener of reasoning with two subtests, Analogic Reasoning and Quantitative Reasoning. Forty-two individuals from the TOD-C standardization and clinical samples took the UNIT-GAT. The UNIT-GAT is intended to be a screener, and interpretation at the full-scale level is most relevant, rather than consideration of the relationship among the individual subtests and the TOD-C scores. Consequently, the correlation between the Full Scale score and the two TOD-C reasoning tests (Picture Analogies [10C], Geometric Analogies [23C]) were of primary interest; moderate correlations were obtained.

### **TOWRE-2**

The TOWRE-2 was taken by 46 individuals in the TOD-C standardization and clinical samples. It assesses reading efficiency in two subtests: Sight Word Efficiency, which requires the examinee to read real words in 45 seconds; and Phonemic Decoding Efficiency, which requires reading nonwords in 45 seconds. These two subtests combine into a single full-scale score. The TOWRE-2 full-scale score had high correlations with the TOD-C rapid word reading tests (Rapid Pseudoword Reading [19C], Rapid Irregular Word Reading [20C]) and a moderate correlation with the Oral Reading Efficiency (12C) test.

### **TOD-E**

For the TOD-E, convergent validity data were collected for three of the same assessments that were used for the TOD-C convergent validity study: Comprehensive Assessment of Spoken Language, Second Edition (CASL-2; Carrow-Woolfolk, 2017); Woodcock-Johnson IV Tests of Achievement (WJ IV ACH; Schrank et al., 2014a); and Comprehensive Test of Phonological Processing, Second Edition (CTOPP-2; Wagner et al., 2013). Table 5.25 presents the demographic characteristics for the TOD-E validation group. Table 5.26 presents the correlations between the TOD-S and TOD-E tests and their corresponding validation tests.

### **CASL-2**

The CASL-2 Receptive Vocabulary test was administered to a sample of 33 individuals from the TOD-E standardization and clinical samples and correlated moderately with TOD-S Picture Vocabulary (1S).

### **WJ IV ACH**

Three of the tests from the WJ IV ACH used to validate the TOD-C tests were also taken by 50 individuals in the TOD-E standardization and clinical samples: Letter-Word Identification, Spelling, and Word Attack. As in the TOD-C sample study, both Letter-Word Identification and Spelling correlated highly with Letter and Word Choice (2S). Letter-Word Identification also correlated highly with Letter and Sight Word Recognition (7E), and Word Attack correlated highly with Sounds and Pseudowords (4E).

### **CTOPP-2**

Four of the subtests from the CTOPP-2 used to validate the TOD-C tests were also taken by 31 individuals from the TOD-E standardization and clinical samples: Elision, Blending Words, Rapid Digit Naming, and Rapid Letter Naming. Elision correlated moderately with Rhyming (5E), while Blending Words demonstrated moderate to high correlations with Early Segmenting (8E) and Letter and Sound Knowledge (9E). The two CTOPP-2 rapid naming tests correlated moderately with the TOD-E Early Rapid Number and Letter Naming (6E) test.

**Table 5.25.** Demographic Characteristics of the TOD-S/TOD-E Convergent Validation Samples

	WJ IV ACH	CASL-2	CTOPP-2
<b>Total sample</b>	50	33	31
<b>Gender</b>			
Male	22	12	17
Female	28	21	14
<b>Parents' educational level</b>			
No high school diploma	6	0	1
High school graduate	9	0	6
Some college	12	10	11
Bachelor's degree or higher	23	23	13
<b>Race/Ethnicity<sup>a</sup></b>			
Asian	8	8	4
Black/African American	9	0	2
White	11	16	11
Other/Multiracial	3	1	2
Hispanic Origin	19	8	12
<b>U.S. geographic region</b>			
Northeast	26	4	0
Midwest	15	9	18
South	8	13	5
West	1	7	8
<b>Age (years)<sup>b</sup></b>			
5	14	3	11
6	12	7	8
7	16	15	10
8–9:3	8	8	2
<b>Disability status</b>			
Clinical	22	7	8
Typical	28	26	23

Note. WJ IV ACH = Woodcock-Johnson IV Tests of Achievement; CASL-2 = Comprehensive Assessment of Spoken Language, Second Edition; CTOPP-2 = Comprehensive Test of Phonological Processing, Second Edition.

<sup>a</sup>Individuals of Hispanic origin are included in the race/ethnicity category under Hispanic Origin; remaining categories include only individuals of non-Hispanic origin.

<sup>b</sup>8-year validation group extends through age 9 years, 3 months.



Table 5.26. Convergent Validation Correlations: TOD-S/TOD-E

TOD-S/TOD-E Test <sup>a</sup>	CASL-2		WJ IV ACH			CTOPP-2		
	Receptive Vocabulary	Letter-Word Identification	Spelling	Word Attack	Elision	Blending Words	Rapid Digit Naming	Rapid Letter Naming
Picture Vocabulary (1S)	.52							
Letter and Word Choice (2S)		.86	.85					
Sounds and Pseudowords (4E)				.90				
Rhyming (5E)					.62			
Early Rapid Number and Letter Naming (6E)							.76	.71
Letter and Sight Word Recognition (7E)		.91						
Early Segmenting (8E)						.67		
Letter and Sound Knowledge (9E)						.80		

Note. *n* varies by test: CASL-2 *n* = 50; CTOPP-2 *n* = 31. CASL-2 = Comprehensive Assessment of Spoken Language, Second Edition; WJ IV ACH = Woodcock-Johnson IV Tests of Achievement; CTOPP-2 = Comprehensive Test of Phonological Processing, Second Edition.

<sup>a</sup>Sample size for Word Reading Fluency and Question Reading Fluency in the TOD-E sample was too small to conduct validation analysis; evidence in the TOD-C sample supports their validity.



## Detection of Skill Weaknesses

The TOD was designed to detect weaknesses in abilities and skills associated with dyslexia and to aid examiners in screening, diagnosing, and planning interventions. In particular, the TOD-S Dyslexia Risk Index (DRI), TOD-C Dyslexia Diagnostic Index (DDI), and TOD-E Early Dyslexia Diagnostic Index (EDDI) were created to differentiate between individuals either having or being at risk for having dyslexia and those with typical reading skills.

Conditional probability analyses (also known as receiver operating characteristic [ROC] curves) were run to determine the capacity of the TOD to detect skill deficits associated with dyslexia at various cutoff values. For these analyses, children diagnosed with a learning disability in reading were compared to the standardization sample of typically developing children. Analyses were obtained from the DRI for the TOD-S, the DDI for the TOD-C, and the EDDI for the TOD-E.

Results indicated that each measure of the TOD risk and diagnostic scores provided statistically significant improvement over chance in detecting dyslexia status: TOD-S DRI score (area under ROC curve = .972,  $p < .001$ ); TOD-C DDI score (area under ROC curve = .989,  $p < .001$ ); TOD-E EDDI score (area under ROC curve = .989,  $p < .001$ ).

Tables 5.27 to 5.29 display the sensitivity and specificity associated with various standard score (SS) values of the TOD. *Sensitivity* refers to a test's capacity to detect true positive cases of the condition in question, i.e., dyslexia. *Specificity* refers to a test's capacity to exclude true negative cases (persons who do not have the condition in question). Betz et al. (2013) recommend providing sensitivity and specificity

results for multiple values so that clinicians can choose a cutoff score that is best suited to their clinical population.

To illustrate, using a cutoff of 80 (one and a third standard deviations below the mean) for the TOD-S DRI yields a sensitivity value of .80 and specificity of .99. In practical terms, this means that 80% of the individuals with clinical diagnoses associated with dyslexia had standard scores less than or equal to 80, whereas 99% of the typically developing children had standard scores greater than 80. Using a very strict guideline for eligibility, such as a standard score of 70 or less, the specificity is also .99 (i.e., only 1% or fewer of typically developing children had standard scores of 70 or less, which is  $\geq 2$  SD below the mean). However, due to the variability inherent in clinical data, only the most severely impaired individuals will be identified as having dyslexia when using such a strict cutoff value (e.g., sensitivity of .40 for the TOD-S and the TOD-C). This finding demonstrates that a cutoff score of 80 provides a reasonable balance between identifying individuals with dyslexia, while not overidentifying those individuals who do not have dyslexia.

These results serve as a reminder that at any level of test score interpretation, there is a risk of under- or overidentifying children who are in need of intervention. Although the TOD provides a measurement of skill difficulties associated with dyslexia, results should not be used in isolation for diagnosis or treatment planning. Instead, these results should be used in concert with other data (e.g., TOD Rating Scales, parent and teacher interview, review of available records, direct observation, and other assessment results, if available).

**Table 5.27.** Conditional Probability Analysis for Detection of Clinical Cases Using the TOD-S Dyslexia Risk Index (DRI) Standard Score

SS cutoff	Sensitivity	Specificity
70	.40	.99
75	.58	.99
80	.80	.99
85	.93	.96
90	.99	.87

*Note.* The analyzed sample included 179 clinically diagnosed children and 1,486 typically developing children.

**Table 5.28.** Conditional Probability Analysis for Detection of Clinical Cases Using the TOD-C Dyslexia Diagnostic Index (DDI) Standard Score

SS cutoff	Sensitivity	Specificity
70	.40	.99
75	.54	.99
80	.78	.97
85	.94	.91
90	.99	.82

*Note.* The analyzed sample included 160 clinically diagnosed children and 1,285 typically developing children.

**Table 5.29.** Conditional Probability Analysis for Detection of Clinical Cases Using the TOD-E Early Dyslexia Diagnostic Index (EDDI) Standard Score

SS cutoff	Sensitivity	Specificity
70	.34	.99
75	.63	.99
80	.80	.99
85	.98	.94
90	.99	.84

*Note.* The analyzed sample included 21 clinically diagnosed children and 249 typically developing children.

## Validity Evidence for Clinical Groups

An important practical aspect of validity is the capacity of test scores to distinguish typically developing individuals from individuals who are expected to perform differently in the measured ability.

In analyzing the clinical groups for the TOD, a randomized, matched control group was drawn from the typical sample, separately for each comparison of group means. Each clinical case was paired with a case of the same age, gender, and parent education level. The means of the two groups were then compared across TOD test, index, and composite scores. Effect sizes are reported to determine whether the group differences are large enough to be considered clinically meaningful. As previously noted, an effect size of 0.2 is considered small, 0.5 is considered medium, and 0.8 is considered large (Cohen, 1992). By convention, an effect size is considered clinically meaningful only if it is medium or larger in magnitude. In this analysis, scores from TOD-S are reported along with the TOD-C and TOD-E samples.

### *TOD-C Child Clinical Sample*

The TOD-C child clinical sample included 511 individuals ages 6–18 years. Chapter 4 describes the sample demographics and diagnostic breakdown. For the clinical discrimination study, this sample was divided into seven groups. Five of these groups were expected to demonstrate differences in TOD scores when compared with a matched typically developing sample: reading learning disability (RLD), language disorder, attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), and intellectual disability (ID) and developmental delay (DD). Two of the groups were expected to show minimal differences in TOD scores: speech disorder and a combined group that included emotional disorders, deaf/hard of hearing, visually impaired, and other health/mental health conditions not accounted for by any other group.

### Reading Learning Disability

The primary clinical group of interest for the TOD consists of 278 individuals diagnosed with dyslexia or a learning disability in reading. (Note that some individuals from this group had comorbid clinical diagnoses and thus are represented in more than one group.) Table 5.30 shows the descriptive statistics

and effect sizes for the comparisons between this clinical group and their corresponding matched control group. The expectation was that the measures of reading and spelling would show large effect sizes of the differences between group means, while the linguistic processing, vocabulary, and reasoning measures would have smaller differences in effect sizes. The results support this expectation.

The effect sizes of the differences between group means for the reading and spelling tests were all large, ranging from 1.05 to 1.50, while those for the linguistic processing, vocabulary, and reasoning tests were medium to large, ranging from 0.31 to 1.11. The effect sizes for the DRI and DDI were also both large, 1.42 and 1.41, respectively, reflecting the validity of the risk and diagnostic scores to differentiate between individuals diagnosed with dyslexia or a learning disability in reading and those who were not. Similarly, effect sizes of the mean differences in the index and composite scores measuring reading and spelling skills were large, ranging from 1.31 to 1.71, while those for index and composite scores measuring other skills were medium to large, ranging from 0.57 to 1.12.

These results provide further validation for the TOD-C by illustrating that the biggest differences in scores between individuals with dyslexia or a learning disability in reading and their matched controls were in reading and spelling; these are the precise skills in which individuals with these diagnoses have the greatest difficulty. Overall, the TOD-C scores distinguish well between individuals who are at risk for having dyslexia or a learning disability in reading and those who are not.

### Language Disorder

Another clinical group of interest for the TOD is individuals diagnosed with a language disorder. Table 5.31 shows the descriptive statistics and effect sizes for the comparisons between this clinical group of 33 individuals and their corresponding matched control group. Because individuals with a developmental language disorder are heterogenous in terms of the manifestation of the disorder, some may have specific difficulty with reading and spelling, whereas others may not. Thus, the expectation was that the tests of reading and spelling would show larger effect sizes between group means than the linguistic processing, vocabulary, and reasoning tests. However,

the differences were not expected to be as large in magnitude as for the RLD sample.

The effect sizes of the differences between group means for the reading and spelling tests were medium to large, ranging from 0.62 to 1.17, while those for the linguistic processing, vocabulary, and reasoning tests were small to large, ranging from 0.20 to 1.00. The effect sizes for the DRI and DDI were also both large, 0.70 and 0.93, respectively. Similarly, effect sizes of the mean differences in the index and composite scores measuring reading and spelling skills were large, ranging from 0.78 to 1.01, while those for index and composite scores measuring other skills were medium to large, ranging from 0.49 to 0.84.

These results support the ability of the TOD tests, indexes, and composites in the skill areas of reading and spelling to distinguish individuals who have developmental language disorders from those who do not. They also distinguish well between the two groups in most other associated skills of linguistic processing, vocabulary, and reasoning.

#### **ADHD**

It is not unusual for individuals with ADHD to have difficulties in reading and spelling due to high comorbidity between these two disorders. However, challenges for individuals with ADHD often extend to other skill areas. Such individuals also often have areas of strength in which they perform quite similarly to their typically developing peers. Because ADHD does not necessarily affect one specific skill, the expectation was to find a range of effect sizes when comparing 118 individuals with a primary diagnosis of ADHD with a matched control group. Table 5.32 shows the descriptive statistics and effect sizes for these comparisons.

The effect sizes of the differences between group means for the reading and spelling tests were medium to large, ranging from 0.53 to 0.90, while those for the linguistic processing, vocabulary, and reasoning tests were small to large, ranging from 0.00 to 0.79. The effect sizes for the DRI and DDI were both large, 0.79 and 1.00, respectively. Effect sizes

of the mean differences in the index and composite scores measuring reading and spelling skills were medium to large, ranging from 0.61 to 0.90. Effect sizes of mean differences of index and composite scores measuring other skills were also medium to large though a bit smaller in magnitude, ranging from 0.42 to 0.81.

Overall, these effect sizes demonstrate the ability of the TOD tests to distinguish individuals with ADHD who have weaknesses in reading and spelling and linguistic processing from those who do not.

#### **Autism Spectrum Disorder**

Individuals who are diagnosed with autism spectrum disorder (ASD) often show difficulties in language and related skills. Table 5.33 shows the descriptive statistics and effect sizes for the 49 individuals diagnosed with ASD and their corresponding matched sample. The results show a similar range of effect sizes across all tests, indexes, and composites. The effect sizes of the mean differences between tests ranged from small to large for the reading and spelling tests (0.31 to 1.02), as well as for the tests of linguistic processing, vocabulary, and reasoning (0.33 to 0.87). Effect sizes for the index and composite scores were mostly in the medium range, though a few were large (0.45 to 0.84).

These results indicate that the TOD is sensitive to the difficulties that are often present for individuals with ASD.

#### **Intellectual Disability and Developmental Delay**

Individuals with a diagnosis of intellectual disability (ID) or developmental delay (DD) generally demonstrate skill deficits across most or all skills measured by the TOD. Thirty-four individuals with a diagnosis of ID or DD were compared with a matched control group. Table 5.34 shows the descriptive statistics and effect sizes for these comparisons. Effect sizes for all tests, indexes, and composites were large, ranging from 0.69 to 1.98. This indicates that the TOD test, index, and composite scores distinguish meaningfully between typically developing individuals and those diagnosed with ID or DD.

**Table 5.30.** TOD-C Child Standard Scores: Descriptive Statistics and Effect Sizes for Individuals With a Reading Learning Disability (RLD) and Matched Control Group

Test/Index/Composite <sup>a</sup>	<i>n</i>	Individuals with RLD		Matched control group		Effect size <sup>b</sup>
		Mean	<i>SD</i>	Mean	<i>SD</i>	
<b>Test</b>						
Picture Vocabulary	268	86.22	20.10	102.36	14.03	0.80
Letter and Word Choice	268	82.81	15.25	101.31	14.04	1.21
Question Reading Fluency	256	84.51	16.71	102.58	13.69	1.08
Phonological Manipulation	276	85.58	16.59	102.94	14.65	1.05
Irregular Word Spelling	277	82.60	15.79	102.05	14.50	1.23
Rapid Letter Naming	274	82.84	16.72	101.41	14.26	1.11
Pseudoword Reading	276	83.19	15.09	102.68	13.49	1.29
Word Pattern Choice	277	92.17	16.33	101.08	15.11	0.55
Word Memory	276	93.39	14.73	101.34	13.94	0.54
Picture Analogies	277	90.60	19.82	102.12	14.16	0.58
Irregular Word Reading	277	80.52	17.65	100.75	14.08	1.15
Oral Reading Efficiency	276	83.84	16.65	101.95	14.18	1.09
Blending	277	91.96	22.32	101.37	15.42	0.42
Segmenting	277	93.85	22.36	100.83	14.95	0.31
Regular Word Spelling	276	83.10	14.80	102.35	14.57	1.30
Silent Reading Efficiency Grades 1–5	155	80.72	19.32	102.24	14.94	1.11
Silent Reading Efficiency Grade 6–Adult	116	85.80	15.84	102.46	13.38	1.05
Rapid Number and Letter Naming	277	84.15	17.24	102.73	14.20	1.08
Letter Memory	277	93.12	15.56	100.77	13.19	0.49
Rapid Pseudoword Reading	246	83.54	12.88	102.86	14.25	1.50
Rapid Irregular Word Reading	260	82.56	14.20	101.70	13.01	1.35
Symbol to Sound Learning	276	93.20	17.63	102.13	14.98	0.51
Listening Vocabulary	277	88.13	18.98	101.99	13.81	0.73
Geometric Analogies	276	88.72	17.33	103.04	14.13	0.83

Note. *N* = 278; some comparisons have smaller *n*s due to missing scores. Means and *SD*s are expressed in standard score units (*M* = 100, *SD* = 15). All pairs of means differ significantly, *p* < .001.

<sup>a</sup>Sample size for Word Reading Fluency test was too small to include.

<sup>b</sup>Effect size (Cohen's *d*) = control group mean minus clinic-referred group mean, divided by pooled standard deviation.

**Table 5.30.** TOD-C Child Standard Scores: Descriptive Statistics and Effect Sizes for Individuals With a Reading Learning Disability (RLD) and Matched Control Group (*continued*)

Test/Index/Composite <sup>a</sup>	<i>n</i>	Individuals with RLD		Matched control group		Effect size <sup>b</sup>
		Mean	<i>SD</i>	Mean	<i>SD</i>	
<b>Index</b>						
Dyslexia Risk Index	265	82.16	14.33	102.46	14.35	1.42
Dyslexia Diagnostic Index	263	79.82	16.29	102.75	14.69	1.41
Linguistic Processing Index	273	83.26	17.01	102.23	14.89	1.12
Reading and Spelling Index	265	80.71	12.97	102.84	14.18	1.71
<b>Composite</b>						
Sight Word Acquisition composite	260	80.82	15.93	101.66	13.57	1.31
Phonics Knowledge composite	246	84.12	12.42	103.17	14.47	1.53
Basic Reading Skills composite	276	81.20	14.05	102.15	14.10	1.49
Decoding Efficiency composite	245	82.04	13.65	102.78	13.96	1.52
Spelling composite	276	81.95	15.53	102.56	14.71	1.33
Reading Fluency composite	264	86.12	13.77	102.96	14.71	1.22
Reading Comprehension Efficiency composite	252	83.61	15.88	102.95	14.32	1.22
Phonological Awareness composite	276	88.97	20.08	101.80	15.37	0.64
Rapid Automated Naming composite	274	81.75	17.24	102.42	14.23	1.20
Auditory Working Memory composite	276	91.84	16.05	101.00	14.57	0.57
Orthographic Processing composite	268	84.54	16.75	101.51	14.59	1.01
Vocabulary composite	268	84.75	21.82	102.38	13.56	0.81
Reasoning composite	276	88.81	18.74	103.23	14.21	0.77
Vocabulary and Reasoning 2 composite	268	87.49	19.12	102.46	14.07	0.78
Vocabulary and Reasoning 4 composite	268	86.37	18.52	103.03	14.20	0.90

Note. *N* = 278; some comparisons have smaller *n*s due to missing scores. Means and *SD*s are expressed in standard score units (*M* = 100, *SD* = 15). All pairs of means differ significantly, *p* < .001.

<sup>a</sup>Sample size for Word Reading Fluency test was too small to include.

<sup>b</sup>Effect size (Cohen's *d*) = control group mean minus clinic-referred group mean, divided by pooled standard deviation.



**Table 5.31.** TOD-C Child Standard Scores: Descriptive Statistics and Effect Sizes for Individuals With a Language Disorder and Matched Control Group

Test/Index/Composite <sup>a</sup>	<i>n</i>	Individuals with a language disorder		Matched control group		Effect size <sup>b</sup>
		Mean	<i>SD</i>	Mean	<i>SD</i>	
<b>Test</b>						
Picture Vocabulary	33	89.82	17.23	98.94	13.05	0.53
Letter and Word Choice	33	87.45	16.15	99.58	14.83	0.75
Question Reading Fluency	30	88.33	16.04	98.28	15.12	0.62
Phonological Manipulation	33	86.70	14.20	100.85	13.50	1.00
Irregular Word Spelling	33	84.94	16.20	97.85	13.52	0.80
Rapid Letter Naming	33	90.09	14.78	101.91	14.53	0.80
Pseudoword Reading	33	87.61	13.96	99.27	12.03	0.84
Word Pattern Choice	33	92.45	13.05	96.76	14.77	0.33
Word Memory	33	92.85	13.22	99.70	14.09	0.52
Picture Analogies	33	92.88	14.57	100.85	15.46	0.55
Irregular Word Reading	33	83.24	16.78	98.00	15.06	0.88
Oral Reading Efficiency	33	82.42	14.55	99.48	14.59	1.17
Blending	33	92.36	22.95	100.15	14.91	0.34
Segmenting	33	92.09	23.34	96.73	13.51	0.20
Regular Word Spelling	33	85.18	16.60	96.94	13.37	0.71
Rapid Number and Letter Naming	33	88.91	15.16	100.67	15.78	0.78
Letter Memory	33	94.12	13.55	100.52	12.54	0.47
Rapid Pseudoword Reading	31	87.61	14.11	99.64	13.73	0.85
Rapid Irregular Word Reading	32	87.53	14.34	99.13	12.74	0.81
Symbol to Sound Learning	33	92.39	11.85	101.33	12.56	0.75
Listening Vocabulary	33	87.91	12.57	95.33	14.20	0.59
Geometric Analogies	33	90.18	15.51	100.97	14.20	0.70

Note. *N* = 33; some comparisons have smaller *ns* due to missing scores. Means and *SDs* are expressed in standard score units (*M* = 100, *SD* = 15). All pairs of means differ significantly, *p* < .001.

<sup>a</sup>Sample sizes were too small to include for Word Reading Fluency, Silent Reading Efficiency Grades 1–5, and Silent Reading Efficiency Grade 6–Adult tests; and for Reading Fluency and Reading Comprehension Efficiency composites.

<sup>b</sup>Effect size (Cohen's *d*) = control group mean minus clinic-referred group mean, divided by pooled standard deviation.

**Table 5.31.** TOD-C Child Standard Scores: Descriptive Statistics and Effect Sizes for Individuals With a Language Disorder and Matched Control Group (*continued*)

Test/Index/Composite <sup>a</sup>	<i>n</i>	Individuals with a language disorder		Matched control group		Effect size <sup>b</sup>
		Mean	<i>SD</i>	Mean	<i>SD</i>	
<b>Index</b>						
Dyslexia Risk Index	33	86.76	16.68	98.42	16.49	0.70
Dyslexia Diagnostic Index	33	83.48	16.51	98.82	15.24	0.93
Linguistic Processing Index	33	85.52	14.66	99.48	14.74	0.95
Reading and Spelling Index	33	85.03	15.27	98.42	14.87	0.88
<b>Composite</b>						
Sight Word Acquisition composite	32	84.41	16.44	99.34	12.93	0.91
Phonics Knowledge composite	31	87.55	13.06	99.36	13.09	0.90
Basic Reading Skills composite	33	84.58	13.79	98.48	13.83	1.01
Decoding Efficiency composite	30	87.23	14.25	99.97	13.20	0.89
Spelling composite	33	84.27	16.85	97.48	13.56	0.78
Phonological Awareness composite	33	88.45	21.21	98.88	14.68	0.49
Rapid Automatized Naming composite	33	88.36	15.73	101.55	15.62	0.84
Auditory Working Memory composite	33	91.85	14.60	99.97	14.75	0.56
Orthographic Processing composite	33	87.24	15.11	97.67	15.45	0.69
Vocabulary composite	33	86.91	16.40	96.70	14.39	0.60
Reasoning composite	33	90.42	15.22	101.12	14.84	0.70
Vocabulary and Reasoning 2 composite	33	89.58	15.92	99.55	15.17	0.63
Vocabulary and Reasoning 4 composite	33	87.55	14.36	98.67	15.23	0.77

Note. *N* = 33; some comparisons have smaller *n*s due to missing scores. Means and *SD*s are expressed in standard score units (*M* = 100, *SD* = 15). All pairs of means differ significantly, *p* < .001.

<sup>a</sup>Sample sizes were too small to include for Word Reading Fluency, Silent Reading Efficiency Grades 1–5, and Silent Reading Efficiency Grade 6–Adult tests; and for Reading Fluency and Reading Comprehension Efficiency composites.

<sup>b</sup>Effect size (Cohen's *d*) = control group mean minus clinic-referred group mean, divided by pooled standard deviation.



**Table 5.32.** TOD-C Child Standard Scores: Descriptive Statistics and Effect Sizes for Individuals With Attention-Deficit/Hyperactivity Disorder (ADHD) and Matched Control Group

Test/Index/Composite <sup>a</sup>	<i>n</i>	Individuals with ADHD		Matched control group		Effect size <sup>b</sup>
		Mean	<i>SD</i>	Mean	<i>SD</i>	
<b>Test</b>						
Picture Vocabulary	112	97.53	13.32	103.50	13.93	0.45
Letter and Word Choice	112	92.88	14.32	102.76	14.10	0.69
Question Reading Fluency	112	94.02	14.79	104.94	14.17	0.74
Phonological Manipulation	118	94.35	13.91	105.14	14.32	0.78
Irregular Word Spelling	118	91.06	15.48	103.31	14.13	0.79
Rapid Letter Naming	117	92.47	14.54	103.99	13.67	0.79
Pseudoword Reading	118	93.62	13.75	102.52	13.10	0.65
Word Pattern Choice	118	96.14	13.86	103.52	14.76	0.53
Word Memory	118	95.04	14.40	102.87	14.37	0.54
Picture Analogies	118	99.49	14.63	103.33	14.56	0.26
Irregular Word Reading	118	93.33	14.92	101.31	13.96	0.53
Oral Reading Efficiency	118	92.99	16.25	103.64	15.29	0.66
Blending	118	96.90	11.13	103.34	14.44	0.58
Segmenting	118	100.04	12.31	100.00	15.40	0.00
Regular Word Spelling	118	91.81	13.99	104.03	14.24	0.87
Silent Reading Efficiency Grades 1–5	60	89.20	16.59	104.15	15.86	0.90
Silent Reading Efficiency Grade 6–Adult	58	96.22	15.60	105.25	14.00	0.58
Rapid Number and Letter Naming	118	94.47	15.81	105.70	14.51	0.71
Letter Memory	118	95.19	13.86	103.22	13.03	0.58
Rapid Pseudoword Reading	118	92.97	14.31	103.70	15.26	0.75
Rapid Irregular Word Reading	118	93.45	14.81	103.23	13.68	0.66
Symbol to Sound Learning	118	99.14	15.34	101.75	14.29	0.17
Listening Vocabulary	118	98.91	13.42	102.99	14.63	0.30
Geometric Analogies	118	94.14	16.09	104.67	13.44	0.65

Note. *N* = 118; some comparisons have smaller *ns* due to missing scores. Means and *SDs* are expressed in standard score units (*M* = 100, *SD* = 15). All pairs of means differ significantly, *p* < .001.

<sup>a</sup>Sample size for Word Reading Fluency test was too small to include.

<sup>b</sup>Effect size (Cohen's *d*) = control group mean minus clinic-referred group mean, divided by pooled standard deviation.

**Table 5.32.** TOD-C Child Standard Scores: Descriptive Statistics and Effect Sizes for Individuals With Attention-Deficit/Hyperactivity Disorder (ADHD) and Matched Control Group (*continued*)

Test/Index/Composite <sup>a</sup>	<i>n</i>	Individuals with ADHD		Matched control group		Effect size <sup>b</sup>
		Mean	<i>SD</i>	Mean	<i>SD</i>	
<b>Index</b>						
Dyslexia Risk Index	112	92.46	15.20	104.50	14.69	0.79
Dyslexia Diagnostic Index	112	90.04	15.10	105.21	15.01	1.00
Linguistic Processing Index	117	91.62	14.48	105.48	15.06	0.96
Reading and Spelling Index	112	91.12	14.71	104.35	14.63	0.90
<b>Composite</b>						
Sight Word Acquisition composite	118	92.84	15.48	102.91	13.60	0.65
Phonics Knowledge composite	118	92.88	14.43	103.62	14.98	0.74
Basic Reading Skills composite	118	92.75	14.03	102.35	13.77	0.68
Decoding Efficiency composite	118	92.69	14.91	104.10	14.61	0.77
Spelling composite	118	91.17	14.88	104.07	14.46	0.87
Reading Fluency composite	112	94.51	14.41	105.64	16.09	0.77
Reading Comprehension Efficiency composite	112	92.74	15.93	105.79	15.20	0.82
Phonological Awareness composite	118	96.10	11.56	103.19	15.41	0.61
Rapid Automatized Naming composite	117	92.95	15.41	105.50	14.43	0.81
Auditory Working Memory composite	118	93.76	15.44	103.38	14.84	0.62
Orthographic Processing composite	112	92.63	14.42	103.94	14.79	0.78
Vocabulary composite	112	98.00	13.27	103.52	13.98	0.42
Reasoning composite	118	96.42	14.87	104.83	14.48	0.57
Vocabulary and Reasoning 2 composite	112	98.05	13.20	103.86	14.31	0.44
Vocabulary and Reasoning 4 composite	112	96.63	13.82	104.67	14.38	0.58

Note. *N* = 118; some comparisons have smaller *n*s due to missing scores. Means and *SD*s are expressed in standard score units (*M* = 100, *SD* = 15). All pairs of means differ significantly, *p* < .001.

<sup>a</sup>Sample size for Word Reading Fluency test was too small to include.

<sup>b</sup>Effect size (Cohen's *d*) = control group mean minus clinic-referred group mean, divided by pooled standard deviation.

**Table 5.33.** TOD-C Child Standard Scores: Descriptive Statistics and Effect Sizes for Individuals With Autism Spectrum Disorder (ASD) and Matched Control Group

Test/Index/Composite <sup>a</sup>	<i>n</i>	Individuals with ASD		Matched control group		Effect size <sup>b</sup>
		Mean	<i>SD</i>	Mean	<i>SD</i>	
<b>Test</b>						
Picture Vocabulary	46	92.83	20.05	102.82	11.90	0.50
Letter and Word Choice	46	91.93	18.53	99.92	12.53	0.43
Question Reading Fluency	40	90.88	17.46	103.76	11.86	0.74
Phonological Manipulation	49	86.84	17.41	102.00	15.23	0.87
Irregular Word Spelling	49	89.71	15.86	100.49	14.92	0.68
Rapid Letter Naming	49	89.10	18.13	99.45	13.02	0.57
Pseudoword Reading	49	92.37	15.68	99.69	12.72	0.47
Word Pattern Choice	49	91.86	13.76	99.24	14.82	0.54
Word Memory	49	96.43	16.68	101.98	15.51	0.33
Picture Analogies	49	93.31	15.33	102.41	17.17	0.59
Irregular Word Reading	49	91.41	18.88	100.16	13.73	0.46
Oral Reading Efficiency	48	93.31	18.70	99.02	15.05	0.31
Blending	49	90.27	21.76	100.88	17.27	0.49
Segmenting	49	89.00	19.08	100.37	15.44	0.60
Regular Word Spelling	49	88.94	16.79	101.33	15.04	0.74
Silent Reading Efficiency Grades 1–5	28	86.75	18.82	102.36	14.63	0.83
Silent Reading Efficiency Grade 6–Adult	20	88.60	14.99	103.88	15.46	1.02
Rapid Number and Letter Naming	49	87.90	17.50	98.37	14.45	0.60
Letter Memory	49	94.29	18.45	102.41	12.21	0.44
Rapid Pseudoword Reading	46	91.46	17.01	100.88	14.19	0.55
Rapid Irregular Word Reading	47	91.04	17.10	98.94	14.19	0.46
Symbol to Sound Learning	49	92.43	14.58	102.39	13.44	0.68
Listening Vocabulary	49	89.33	16.09	101.80	13.06	0.77
Geometric Analogies	49	93.86	16.92	100.27	14.67	0.38

Note. *N* = 49; some comparisons have smaller *n*s due to missing scores. Means and *SD*s are expressed in standard score units (*M* = 100, *SD* = 15). All pairs of means differ significantly, *p* < .001.

<sup>a</sup>Sample sizes for Word Reading Fluency test and for Reading Fluency and Reading Comprehension Efficiency composites were too small to include.

<sup>b</sup>Effect size (Cohen's *d*) = control group mean minus clinic-referred group mean, divided by pooled standard deviation.

**Table 5.33.** TOD-C Child Standard Scores: Descriptive Statistics and Effect Sizes for Individuals With Autism Spectrum Disorder (ASD) and Matched Control Group (*continued*)

Test/Index/Composite <sup>a</sup>	<i>n</i>	Individuals with ASD		Matched control group		Effect size <sup>b</sup>
		Mean	<i>SD</i>	Mean	<i>SD</i>	
<b>Index</b>						
Dyslexia Risk Index	45	89.69	18.06	101.94	12.55	0.68
Dyslexia Diagnostic Index	45	87.62	17.20	101.35	14.18	0.80
Linguistic Processing Index	49	86.41	16.97	100.69	16.07	0.84
Reading and Spelling Index	45	89.93	16.66	101.35	12.92	0.69
<b>Composite</b>						
Sight Word Acquisition composite	47	90.43	18.45	99.71	13.98	0.50
Phonics Knowledge composite	46	92.46	16.51	100.45	14.17	0.48
Basic Reading Skills composite	49	91.49	17.57	99.94	13.22	0.48
Decoding Efficiency composite	45	90.64	18.02	100.06	14.83	0.52
Spelling composite	49	88.94	16.90	101.20	15.32	0.73
Phonological Awareness composite	49	86.18	19.44	100.96	16.34	0.76
Rapid Automatized Naming composite	49	87.08	19.05	98.88	13.78	0.62
Auditory Working Memory composite	49	93.73	19.01	102.31	15.43	0.45
Orthographic Processing composite	46	89.50	18.11	99.55	14.46	0.55
Vocabulary composite	46	88.72	20.30	102.63	12.26	0.69
Reasoning composite	49	92.94	17.24	101.67	15.98	0.51
Vocabulary and Reasoning 2 composite	46	91.74	18.01	103.00	13.97	0.63
Vocabulary and Reasoning 4 composite	46	90.09	17.85	102.27	14.25	0.68

Note. *N* = 49; some comparisons have smaller *n*s due to missing scores. Means and *SD*s are expressed in standard score units (*M* = 100, *SD* = 15). All pairs of means differ significantly, *p* < .001.

<sup>a</sup>Sample sizes for Word Reading Fluency test and for Reading Fluency and Reading Comprehension Efficiency composites were too small to include.

<sup>b</sup>Effect size (Cohen's *d*) = control group mean minus clinic-referred group mean, divided by pooled standard deviation.

**Table 5.34.** TOD-C Child Standard Scores: Descriptive Statistics and Effect Sizes for Individuals With Intellectual Disability (ID) or Developmental Delay (DD) and Matched Control Group

Test/Index/Composite <sup>a</sup>	n	Individuals with ID or DD		Matched control group		Effect size <sup>b</sup>
		Mean	SD	Mean	SD	
<b>Test</b>						
Picture Vocabulary	33	80.58	17.26	97.18	15.33	0.96
Letter and Word Choice	33	75.94	14.87	96.38	14.66	1.37
Question Reading Fluency	32	72.78	14.71	96.76	16.31	1.63
Phonological Manipulation	34	69.65	16.27	93.29	17.74	1.45
Irregular Word Spelling	34	70.79	16.43	93.50	17.51	1.38
Rapid Letter Naming	34	71.85	14.48	93.94	18.72	1.53
Pseudoword Reading	34	73.38	12.54	92.74	15.37	1.54
Word Pattern Choice	34	82.24	11.96	100.41	14.76	1.52
Word Memory	34	80.94	18.40	94.65	14.15	0.75
Picture Analogies	34	80.56	14.12	94.32	15.51	0.97
Irregular Word Reading	34	71.03	18.42	93.79	18.51	1.24
Oral Reading Efficiency	27	72.78	13.80	94.39	13.31	1.57
Blending	34	76.21	22.90	92.03	18.93	0.69
Segmenting	34	75.06	21.44	90.82	13.98	0.74
Regular Word Spelling	34	72.03	15.56	93.00	15.28	1.35
Rapid Number and Letter Naming	34	76.03	13.42	95.41	17.88	1.44
Letter Memory	34	77.85	18.96	97.79	13.16	1.05
Rapid Pseudoword Reading	33	72.12	11.98	94.21	16.61	1.84
Rapid Irregular Word Reading	34	72.65	14.24	96.61	15.21	1.68
Symbol to Sound Learning	34	79.26	15.61	97.56	17.07	1.17
Listening Vocabulary	34	73.50	16.41	89.35	13.56	0.97
Geometric Analogies	34	80.38	12.67	98.15	13.69	1.40

Note.  $N = 34$ ; some comparisons have smaller  $n$ s due to missing scores. Means and SDs are expressed in standard score units ( $M = 100, SD = 15$ ). All pairs of means differ significantly,  $p < .001$ .

<sup>a</sup>Sample sizes were too small to include for Word Reading Fluency, Silent Reading Efficiency Grades 1–5, and Silent Reading Efficiency Grade 6–Adult tests; and for Reading Fluency and Reading Comprehension Efficiency composites.

<sup>b</sup>Effect size (Cohen's  $d$ ) = control group mean minus clinic-referred group mean, divided by pooled standard deviation.

**Table 5.34.** TOD-C Child Standard Scores: Descriptive Statistics and Effect Sizes for Individuals With Intellectual Disability (ID) or Developmental Delay (DD) and Matched Control Group (*continued*)

Test/Index/Composite <sup>a</sup>	<i>n</i>	Individuals with ID or DD		Matched control group		Effect size <sup>b</sup>
		Mean	<i>SD</i>	Mean	<i>SD</i>	
<b>Index</b>						
Dyslexia Risk Index	32	72.31	14.44	96.30	15.76	1.66
Dyslexia Diagnostic Index	32	63.31	16.50	93.12	17.61	1.81
Linguistic Processing Index	34	65.18	15.12	93.12	17.60	1.85
Reading and Spelling Index	32	70.56	13.18	94.09	16.03	1.78
<b>Composite</b>						
Sight Word Acquisition composite	34	68.59	18.18	95.55	17.34	1.48
Phonics Knowledge composite	33	73.70	10.34	93.44	16.14	1.91
Basic Reading Skills composite	34	72.15	13.14	92.88	16.54	1.58
Decoding Efficiency composite	33	69.79	12.92	95.42	16.65	1.98
Spelling composite	34	69.59	16.98	93.03	16.37	1.38
Phonological Awareness composite	34	68.50	19.75	89.79	18.67	1.08
Rapid Automatized Naming composite	34	70.82	13.80	93.91	18.46	1.67
Auditory Working Memory composite	34	77.06	18.54	95.24	15.31	0.98
Orthographic Processing composite	33	72.82	15.65	98.00	14.42	1.61
Vocabulary composite	33	72.67	18.62	92.32	13.97	1.06
Reasoning composite	34	78.35	12.96	95.74	13.63	1.34
Vocabulary and Reasoning 2 composite	33	77.67	13.91	94.68	14.64	1.22
Vocabulary and Reasoning 4 composite	33	74.48	13.93	93.15	13.81	1.34

Note. *N* = 34; some comparisons have smaller *n*s due to missing scores. Means and *SD*s are expressed in standard score units (*M* = 100, *SD* = 15). All pairs of means differ significantly, *p* < .001.

<sup>a</sup>Sample sizes were too small to include for Word Reading Fluency, Silent Reading Efficiency Grades 1–5, and Silent Reading Efficiency Grade 6–Adult tests; and for Reading Fluency and Reading Comprehension Efficiency composites.

<sup>b</sup>Effect size (Cohen's *d*) = control group mean minus clinic-referred group mean, divided by pooled standard deviation.

### Other Clinical Groups

Some individuals in the TOD-C child clinical sample had primary diagnoses that were unlikely to show meaningful differences based on the TOD. One such group is individuals with a primary diagnosis of a speech or articulation disorder ( $n = 62$ ). The TOD tests were not designed to be sensitive to speech or articulation problems; however, because there is a relationship between language and speech, small differences would not be unexpected. Results show that the effect sizes of the differences in test means between individuals with a speech or articulation disorder and their matched controls ranged from small to medium (0.02 to 0.62), though most were small. The effect sizes of the mean differences between indexes and composites also ranged from small to medium (0.02 to 0.59), though most were small as well.

The other group of individuals unlikely to show differences on the TOD are those diagnosed with emotional/mood disorder, deaf/hard of hearing, visual impairment, or other health/mental health impairment. Comparisons of these 53 individuals with a matched control group revealed almost all small effect sizes, with test effect sizes ranging from 0.01 to 0.42 and index and composite effect sizes ranging from 0.02 to 0.43.

These results contribute validation evidence by demonstrating smaller effect sizes for clinical groups whose primary deficits are not related to reading and spelling.

### TOD-C Adult Clinical Sample

The TOD-C adult clinical sample was composed primarily of individuals unlikely to have specific challenges on skills measured by the TOD. This includes individuals diagnosed with emotional/mood disorder, deaf/hard of hearing, visual impairment, or other health/mental health impairment. In adults, this also includes individuals with ADHD. Unlike children with the diagnosis, adults typically have had intervention and/or developed compensating strategies that make their skill levels somewhat comparable to adults without an ADHD diagnosis.

The adult Learning Disability in Reading sample was composed of only 16 individuals and thus was too small for a reliable analysis. That said, this group showed a DRI mean of 90.81 and DDI mean of 93.06, which are in the expected direction.

### TOD-E Clinical Sample

#### Reading Learning Disability

The primary clinical group of interest for the TOD-E consisted of 31 individuals diagnosed with dyslexia or a learning disability in reading. (Note that some individuals from this group had comorbid clinical diagnoses and thus are represented in more than one group.) Table 5.35 shows the descriptive statistics and effect sizes for the comparisons between this clinical group and their corresponding matched control group. The expectation was that the measures of reading and spelling would show large effect sizes of the differences between group means, while the effect sizes for the measures of linguistic processing and vocabulary were expected to be lower. Results support this expectation.

The effect sizes of the differences between group means for the reading and spelling tests are all large, ranging from 0.91 to 1.21, and those of the linguistic processing, vocabulary, and reasoning tests were medium to large, ranging from 0.59 to 1.23. The effect sizes for the DRI and EDDI were also both large, 0.83 and 1.20, respectively, reflecting the validity of the risk and diagnostic scores to differentiate between individuals diagnosed with dyslexia or a learning disability in reading and those who were not. Effect sizes of the mean differences in the index and composite scores measuring reading and spelling skills were also large, ranging from 1.03 to 1.38, as were those of scores measuring other related skills, ranging from 0.94 to 1.14.

These results provide further validation for the TOD by illustrating that the TOD-E scores distinguish well between individuals who have dyslexia or a learning disability in reading and those who do not.

**Table 5.35.** TOD-E Child Standard Scores: Descriptive Statistics and Effect Sizes for Individuals With a Reading Learning Disability (RLD) and Matched Control Group

Test/Index/Composite <sup>a</sup>	<i>n</i>	Individuals with RLD		Matched control group		Effect size <sup>b</sup>
		Mean	<i>SD</i>	Mean	<i>SD</i>	
<b>Test</b>						
Picture Vocabulary	30	82.27	22.37	98.90	12.14	0.74
Letter and Word Choice	30	82.90	14.05	95.68	14.87	0.91
Sounds and Pseudowords	31	83.81	15.01	99.90	12.85	1.07
Rhyming	31	86.81	11.99	100.16	12.65	1.11
Early Rapid Number and Letter Naming	31	81.16	13.30	97.58	16.08	1.23
Letter and Sight Word Recognition	30	84.77	13.61	97.32	12.90	0.92
Early Segmenting	31	89.81	16.21	99.32	11.46	0.59
Letter and Sound Knowledge	30	80.30	13.45	96.58	11.67	1.21
<b>Index</b>						
Dyslexia Risk Index standard score	29	81.38	16.23	94.81	15.61	0.83
Early Dyslexia Diagnostic Index	27	80.78	13.68	97.23	14.52	1.20
Early Linguistic Processing Index	31	81.68	15.19	99.06	14.06	1.14
Early Reading and Spelling Index	27	80.63	13.82	96.58	14.50	1.15
<b>Composite</b>						
Early Sight Word Acquisition composite	29	82.28	13.93	96.65	15.09	1.03
Early Phonics Knowledge composite	30	80.20	13.20	98.35	13.20	1.38
Early Basic Reading Skills composite	29	80.59	11.85	96.77	14.04	1.37
Early Phonological Awareness composite	31	85.32	15.88	100.32	12.76	0.94

Note. *N* = 31; some comparisons have smaller *n*s due to missing scores. Means and *SD*s are expressed in standard score units (*M* = 100, *SD* = 15). All pairs of means differ significantly, *p* < .001.

<sup>a</sup>Sample sizes for Word Reading Fluency and Question Reading Fluency tests were too small to include.

<sup>b</sup>Effect size (Cohen's *d*) = control group mean minus clinic-referred group mean, divided by pooled standard deviation.

### Combined Clinical Group

Due to the smaller number of individuals with a clinical diagnosis in the TOD-E sample, all clinical diagnoses of interest were collapsed into a single group. This group of 80 individuals had diagnoses of developmental delay, intellectual disability, language disorder, autism spectrum disorder, and attention-deficit/hyperactivity disorder. Table 5.36 shows the descriptive statistics and effect sizes for the comparisons between this clinical group and their corresponding matched control group. Because individuals with these diagnoses are likely to have skill difficulties across all areas measured by the TOD (with the possible exception of those with a language

disability), there wasn't an expectation of a difference between tests, indexes, or composites of reading and spelling compared with those of linguistic processing or vocabulary. However, effect sizes of the differences between the means of the clinical group compared with the matched control were medium to large (mostly large), ranging from 0.46 to 1.08.

This indicates that the TOD-E test, index, and composite scores distinguish meaningfully between typically developing individuals and individuals with one or more of the following diagnoses: developmental delay, intellectual disability, language disorder, autism spectrum disorder, and attention-deficit/hyperactivity disorder.



## Other Clinical Groups

Twenty-nine individuals in the TOD-E clinical sample had a primary diagnosis of a speech or articulation disorder that was unlikely to show meaningful differences based on the TOD. Comparing their mean differences with a matched control group yielded small to medium effect sizes. Test effect sizes ranged from 0.01 to 0.39, and index and

composite effect sizes ranged from 0.00 to 0.44.

The TOD tests were not designed to be sensitive to speech or articulation problems; however, because there is a relationship between language and speech, particularly at younger ages, these differences are not unexpected. This absence of large effect sizes in a group not expected to differ meaningfully on the skills measured by the TOD-E tests contributes validation evidence for the TOD.

**Table 5.36.** TOD-E Child Standard Scores: Descriptive Statistics and Effect Sizes for Individuals With Developmental Delay, Intellectual Disability, Language Disorder, Autism Spectrum Disorder, Attention-Deficit/Hyperactivity Disorder, and Matched Control Group

Test/Index/Composite	<i>n</i>	Clinical group		Matched control group		Effect size <sup>a</sup>
		Mean	<i>SD</i>	Mean	<i>SD</i>	
<b>Test</b>						
Picture Vocabulary	80	83.68	17.93	97.73	16.87	0.78
Letter and Word Choice	80	88.50	15.37	98.13	15.10	0.63
Word Reading Fluency	23	85.83	18.53	104.75	19.42	1.02
Question Reading Fluency	55	92.04	15.62	99.16	14.93	0.46
Sounds and Pseudowords	80	87.91	15.57	101.10	14.15	0.85
Rhyming	80	86.66	11.66	99.30	14.67	1.08
Early Rapid Number and Letter Naming	80	87.65	15.20	97.40	16.59	0.64
Letter and Sight Word Recognition	80	89.28	15.30	100.13	13.71	0.71
Early Segmenting	78	89.03	12.87	96.54	15.34	0.58
Letter and Sound Knowledge	80	86.64	15.45	97.90	13.65	0.73
<b>Index</b>						
Dyslexia Risk Index standard score	78	88.65	16.00	99.69	16.61	0.69
Early Dyslexia Diagnostic Index	76	85.66	14.21	99.73	15.56	0.99
Early Linguistic Processing Index	78	84.08	13.63	97.41	16.14	0.98
Early Reading and Spelling Index	78	87.28	15.64	100.72	15.47	0.86
<b>Composite</b>						
Early Sight Word Acquisition composite	80	88.36	15.28	100.10	15.51	0.77
Early Phonics Knowledge composite	80	85.54	16.99	99.84	15.11	0.84
Early Basic Reading Skills composite	80	86.55	15.97	99.39	15.12	0.80
Early Phonological Awareness composite	78	85.19	12.99	97.78	15.54	0.97

Note. *N* = 80; some comparisons have smaller *n*s due to missing scores. Means and *SD*s are expressed in standard score units (*M* = 100, *SD* = 15). All pairs of means differ significantly, *p* < .001.

<sup>a</sup>Effect size (Cohen's *d*) = control group mean minus clinic-referred group mean, divided by pooled standard deviation.

## Predictive Validity

### *DDI Predictive Validity*

The predictive power of the TOD-C Dyslexia Diagnostic Index (DDI) was evaluated using binary logistic regression analyses with two groups of 261 students each: 1) students with dyslexia, and 2) a matched control group. Two separate regression analyses were conducted. The first used four TOD tests that operationalize the Simple View of Reading (SVR). SVR posits that Decoding  $\times$  Linguistic Comprehension = Reading Comprehension (Gough & Tunmer, 1986); numerous studies have supported the robustness of the SVR in explaining reading comprehension (Hoover & Tunmer, 2021). Irregular Word Reading (11C) and Pseudoword Reading (7C) were operationalized as measures of decoding, and Picture Vocabulary (1S) and Listening Vocabulary (22C) as measures of listening comprehension. Students with dyslexia were 3.44 times more likely to be predicted as having dyslexia than students without dyslexia when the SVR scores alone were included in the model. Students with dyslexia were 9.29 times more likely to be predicted as having dyslexia than students without dyslexia when TOD-C DDI scores alone were included in the model (Castleman et al., 2023). This large increase in predictive power from the SVR to the DDI illustrates the robust ability of the DDI to accurately predict dyslexia.

### *TOD Rating Scale Predictive Validity*

The TOD Rating Scales were developed to provide another means of gathering information in a comprehensive TOD evaluation, but they can also serve independently to predict the likelihood of dyslexia. Note that the expectation is that correlations between the TOD direct assessment tests and the Rating Scales will be negative because they are scored in opposing directions (e.g., the higher the Rating Scale score, the greater the difficulty for the individual being rated).

### **TOD-C Rating Scales**

Intercorrelation coefficients between the Rating Scales were presented as evidence of cross-form consistency earlier in this chapter in the Reliability section (see Tables 5.6 and 5.9). The correlations reported in Table 5.37 are between ratings completed

on the same individual and thus are higher than those reported in the Reliability section. The sample consisted of individuals with a diagnosis of reading disability who had all three Rating Scales (Parent/Caregiver, Teacher, and Self-Rating) completed and a matched sample of individuals from the standardization sample (for use in logistic regression analyses, described in the next paragraph). As shown in Table 5.37, the correlations between the three Rating Scales are .77, .77, and .81 for the Self-Rating to Parent/Caregiver, Self-Rating to Teacher, and Teacher to Parent/Caregiver comparisons, respectively. The correlations between the Rating Scale and the DRI and DDI scores were all moderate to large, ranging from  $-.64$  to  $-.71$  (and significant at  $p < .001$ ).

Logistic regression analyses were conducted to determine the ability of the TOD-C Rating Scales to detect clinically significant weaknesses in skills associated with dyslexia by predicting membership in the group of individuals diagnosed with reading disability or in the matched control group. Results indicated that each of the TOD-C Rating Scales provides statistically significant improvement over chance in detecting reading disability status. The percentages of correct diagnostic decisions were 77%, 82%, and 83% for the Parent/Caregiver, Teacher, and Self-Rating scales, respectively. Thus, the Rating Scales are credible predictors of students who have a learning disability in reading, and consequently most likely those who have dyslexia.

### **TOD-E Rating Scales**

The TOD-E sample included only Parent/Caregiver and Teacher Rating Scales. The results in Table 5.38 are based on individuals for whom both Rating Scales were completed. The intercorrelation coefficient between the Rating Scales was .75. Correlation coefficients between the Rating Scales and the TOD-S DRI and TOD-E EDDI were moderate, ranging from  $-.33$  to  $-.55$  (and significant at  $p < .001$ ). Although these correlations are smaller in magnitude than for the TOD-C, they still demonstrate a relationship between the Rating Scales and the DRI and EDDI. Because of the small number of students identified with a reading disability in the TOD-E sample, no logistic regression analyses like those conducted for the TOD-C were conducted for the TOD-E.

**Table 5.37.** Correlations Between TOD Rating Scales and Dyslexia Risk and Diagnostic Index Standard Scores: TOD-C

	Parent/Caregiver Rating Scale	Teacher Rating Scale	Self-Rating Scale
Parent/Caregiver Rating Scale	—		
Teacher Rating Scale	.81	—	
Self-Rating Scale	.77	.77	—
DRI	-.71	-.65	-.70
DDI	-.69	-.64	-.65

Note.  $N = 66$ . Correlations are based on Rating Scales completed for the same individual. **DRI** = Dyslexia Risk Index; **DDI** = Dyslexia Diagnostic Index.

**Table 5.38.** Correlations Between TOD Rating Scales and Dyslexia Risk and Diagnostic Index Standard Scores: TOD-E

	Parent/Caregiver Rating Scale	Teacher Rating Scale
Parent/Caregiver Rating Scale	—	
Teacher Rating Scale	.75	—
DRI	-.33	-.40
EDDI	-.51	-.55

Note.  $N = 85$ . Correlations are based on Rating Scales completed for the same individual. **DRI** = Dyslexia Risk Index; **EDDI** = Early Dyslexia Diagnostic Index.

## Summary

This chapter described the psychometric research undertaken to support the publication of the TOD. Reliability was examined from several perspectives, and the test, index, and composite scores performed well based on internal consistency and test–retest reliability analyses. The Rating Scales showed good internal consistency as well as cross-form consistency and validity. A confirmatory factor analysis showed acceptable fit with the theoretical model upon which the TOD was based. Similarly, the TOD tests correlate in expected ways with other tests of similar constructs, thereby yielding evidence of

convergent validity. Finally, the TOD Dyslexia Risk and Diagnostic Indexes distinguish typically developing individuals from those with a reading disability. Treatment outcome research is needed to expand the range of validity evidence for the TOD. Such research should include studies that assess individuals with language disorders and other related disabilities, before and after intervention. These studies will help to validate the TOD as an integral component of evidence-based assessment and intervention planning for individuals with dyslexia.

# Glossary of Terms

---

- alphabetic principle:** the basic understanding that spoken language is made up of speech sounds (phonemes) that can be represented by a letter or letter string (grapheme)
- associative memory:** recall of the connection between two elements, such as letter names and speech sounds
- automaticity:** the ability to recognize words quickly
- connected text:** text that can be read continuously as opposed to word lists
- decodable text:** reading material that includes words with regular sound–symbol correspondences and that is used to practice the application of common phonic elements
- dyslexia:** a neurological disorder that causes a marked impairment in the development of basic reading skills, reading rate, and spelling
- fluency:** the ability to read a text accurately, quickly, and with appropriate expression
- grapheme:** the letter or letter combination that represents a single speech sound (e.g., the *l* in *lap*, the *tch* in *catch*)
- lexical:** relating to the words or vocabulary of a language
- orthographic mapping:** the process of assigning individual speech sounds to the letters that represent those sounds; this process bonds the spelling, pronunciation, and meaning of a specific word in memory and explains how children learn to read sight words
- orthography:** how a language is represented in writing, including the spelling patterns and rules for punctuation and capitalization
- paired-associate learning (PAL):** learning and recalling the associations between two stimuli, such as a symbol and a letter or word
- phoneme:** an individual speech sound (e.g., *cat* has three phonemes: /k/ /ă/ /t/)
- phoneme–grapheme correspondence:** the associations between the speech sounds (phonemes) and the letters representing those sounds (graphemes)
- phonemic awareness:** hearing and using individual speech sounds in words; it includes activities such as combining sounds to read a written word (e.g., putting together the sounds /b/, /ă/, and /g/ to form the word *bag*) or pulling apart the sounds to spell a word
- phonemic manipulation:** tasks that involve altering the order of sounds in a spoken or written word
- phonics:** an instructional reading method for teaching students the relationships between the individual speech sounds and the letter or letters that represent these sounds and how to apply these sound–symbol correspondences to reading and spelling
- phonological awareness:** the umbrella term that encompasses a broad range of tasks that involve understanding and using word parts and speech sounds (e.g., rhyming words, combining the two parts of compound words, counting the number of syllables within words, counting phonemes)
- phonology:** the rule system that governs the relationships among the speech sounds of a language
- prosody:** a component of fluency that includes the patterns of stress and intonation in a language
- receptive vocabulary:** the words that an individual can understand when spoken or read

**segment:** to break apart compound words, syllables, or phonemes of words

**sight word:** any word that a reader recognizes instantly without needing to use decoding strategies

**sound or phonetic spelling:** the words are spelled the way they sound even though the correct letter combinations may not be used