

Technical Adequacy of Acadience Reading Diagnostic CFOL

This document is reprinted from the
Acadience Reading Diagnostic: Comprehension, Fluency, &
Oral Language Assessment Manual Chapter 12.

Kelly A. Powell-Smith

Ruth A. Kaminski

Roland H. Good

with

Mary Abbott

Katherine Bravo Aguayo

Elizabeth N. Dewey

Rachael J. Latimer

Stephanie Stollar

Amy N. Warnock

acadience® reading diagnostic

Suggested Citation:

Powell-Smith, K. A., Kaminski, R. A., Good, R. H., Abbott, M., Bravo Aguayo, K., Dewey, E. N., Latimer, R. J., Stollar, S., & Warnock, A. N., (2023). *Technical Adequacy of Acadience Reading Diagnostic CFOL*. Acadience Learning Inc. www.acadiencelearning.org (Preprinted from *Acadience Reading Diagnostic: Comprehension, Fluency, & Oral Language assessment manual*, pp. 111–143 by K. A. Powell-Smith, R. A. Kaminski, R. H. Good, M. Abbott, K. Bravo Aguayo, E. N. Dewey, R. J. Latimer, S. Stollar, & A. N. Warnock, 2021, Voyager Sopris Learning.

Technical Adequacy of Acadience Reading Diagnostic CFOL

This chapter describes research on the technical adequacy of the Acadience Reading Diagnostic CFOL measures. Three primary studies have been conducted on CFOL. First, we will describe a small pilot study of the measures. Second, we will describe a more extensive piloting of the measures (Pilot-2). Finally, we will describe a validation study which included a factor analysis and item analyses conducted on the measures. While brief overviews of the pilot studies will be provided, we will focus this chapter primarily on the validation study that provides technical adequacy data on the final measures.

CFOL Pilot Study

The purposes of the pilot study were to: (a) examine the scope and sequence of skills; (b) determine if the test directions and procedures functioned in such a way that students understood the tasks; (c) examine the starting points and discontinue rules; (d) determine if scoring rules were appropriate (e.g., functional for assessors); and (e) determine if prompting rules were appropriate (i.e., functional for assessors and served their intended purpose). The pilot study was primarily descriptive and qualitative in nature.

One local school participated in the pilot study. The school had about 400 children in grades K–5. School demographic data indicated that the student to teacher ratio was 23:1; about 53% of the students at the school participated in the federal free and reduced price lunch program; and the student body was approximately 2% American Indian/Alaska Native, 2% Asian/Pacific Islander, 2% Black, 12% Hispanic, and 79% White. A sample of 76 students from this school participated in the piloting of the measures, including 14 students from kindergarten and first grade and 16 students from each grade in second through fourth grades. Trained personnel from Acadience Learning collected all pilot study data. Prior to data collection, parent consent and student assent were obtained.

Five data collectors assessed participating students on designated sections of the CFOL test materials and provided detailed qualitative feedback. In some cases, notes were written directly on assessment materials. Meetings were held with assessors after the testing was completed to discuss the experience and their feedback about what worked well and what might be changed. The qualitative feedback was recorded in an electronic spreadsheet and examined for patterns. Finally, after the pilot study, the materials were revised for the next round of research based upon feedback from our expert reviewer (see Chapter 11) in addition to the feedback from the assessors who participated in the pilot study.

Pilot-2 Study

The Pilot-2 study was a much more extensive piloting of the measures, including an initial examination of the technical adequacy, with a larger sample of students. The purposes of the Pilot-2 study were to examine the (a) accuracy of the scope of sequence of skills, including item difficulty; (b) appropriateness of selected starting points and discontinue rules; (c) relation between Acadience Reading measures and Acadience Reading Diagnostic CFOL tasks; and (d) utility and functionality of the CFOL measures via assessor feedback.

Two school districts in two states participated in Pilot-2. Both sites were located in the Midwest Census Region of the United States. There were six K–6 grade schools in one state/district and one pre-K to grade 4 school in the other state/district that participated. The school sizes ranged from 291 to 574 (median = 412) students. The student to teacher ratio ranged from 13:1 to 18:1 (median = 16:1). The percentage of students in these schools who participated in the federal free and reduced-price lunch program ranged from 19% to 53% (median = 44%). The racial and ethnic composition in these schools varied, but on average there were less than 1% American Indian/Alaska Native, 3% Asian/Pacific Islander, 7% Black, 23% Hispanic, 60% White, and 4% multiracial. From these schools, a total of 762 students across grades K to 4 participated in the Pilot-2 study. About 150 students per grade participated. Specifically, 149 students in each of kindergarten and first grade, 161 second-grade students, 152 third-grade students, and 151 fourth-grade students participated in the study.

All data were collected by trained personnel. Prior to data collection, parental consent and student assent were obtained. Students were administered portions of CFOL appropriate for their grade level, but no student was administered the entire test to reduce testing time. Data analysis for the Pilot-2 study focused on descriptive statistics, item-response analysis, and correlations. Descriptive statistics helped to inform decisions about the scope and sequence, tasks appropriate for each grade, as well as decisions about items. Correlations with Acadience Reading also helped to inform decisions about the structure of the measure.

The Pilot-2 results indicated that, in general, the starting points and discontinue rules appeared to work well. However, the routing rules for the comprehension tasks suggested the starting point was too challenging, in particular for younger students. Regarding the scope and sequence, we examined scores in each test section by grade level. Some sections did not show growth as grade level increased. Other sections showed ceiling effects at upper grades (e.g., some of the story coherence/text structure tasks). Item-level results were used when deciding to move or delete items (e.g., items where no student responded correctly or all students responded correctly were removed or re-worked). Results of examining the correlations between CFOL tasks and Acadience Reading measures were promising with several correlations ranging from moderate to strong. Finally, a sample of assessors responded to an electronic questionnaire administered via SurveyMonkey. There were 15 items and 16 or 17 assessors responded to each item. Items were rated using a Likert-type scale with 1 = Strongly Disagree, 2 = Disagree, 3 = Somewhat Disagree, 4 = Somewhat Agree, 5 = Agree, 6 = Strongly Agree. Responses were by and large very positive, ranging from a mean of 4.13 (SD = 1.02) in response to “All items included within the measure were appropriate” (e.g., all words/passages seemed at the appropriate grade level assigned to them) to 5.29 (SD = 0.69) for “Overall, the measures would be beneficial for planning reading instruction for struggling readers.”

CFOL Validation Study

The purpose of this study was to examine further the use and the utility of brief diagnostic measures of reading comprehension skills linked to Acadience Reading. Specifically, the validation study (a) examined the procedural reliability of the Acadience Reading Diagnostic CFOL measures; (b) examined the appropriateness of the order of items within each task, as well as across sections; (c) confirmed appropriate discontinue rules for the measure; (d) examined the relation of the Acadience Reading Diagnostic CFOL measures to Acadience Reading benchmark data; (e) examined the factor structure of CFOL; and (f) examined consumer satisfaction with the Acadience Reading Diagnostic CFOL measures.

Participants. The participants in this study included four public schools from four states across three major regions of the United States as designated by the U.S. Census Bureau. Demographic data on participating school sites drawn from the National Center for Educational Statistics (NCES) is found in Table 12.1. Three of the participating schools were located in remote rural areas, and one was located in a large suburb. The schools ranged in the size of the student population served from 219 to 487 students, and the grade levels served ranged from pre-kindergarten to sixth. Student to teacher ratios ranged from 11:1 to 18:1.

Three of the four participating schools were Title 1 eligible. Across the participating schools, the percent of students eligible for the federal free and reduced price lunch program ranged from 52% to 97%, with only three of the schools reporting data on this demographic characteristic. The student population in the schools was on average about 50% female (range = 48%–53%). The race and ethnicity composition of the schools ranged as follows: American Indian (0–14%), Asian (0–3%), Black (<1–95%), Hispanic (2–93%), White (<1–91%), and multiracial (0–3%).

Table 12.1
School Demographic Characteristics

Locale	School Number			
	1 Rural: Remote	2 Rural: Remote	3 Suburb: Large	4 Rural: Remote
Grades Taught	PK–5	K–6	K–5	PK–5
Total Students	351	487	219	441
Student/Teacher Ratio	16:1	11:1	18:1	16:1
Title 1 Eligible	-	Yes ¹	Yes ¹	Yes ¹
Free/Reduced Lunch	97%	52%	-	70%
Percent Female	49%	48%	53%	51%
Student Ethnicity				
Am. Indian	0%	<1%	0%	14%
Asian	<1%	0%	3%	0%
Black	95%	3%	<1%	3%
Hispanic	4%	2%	93%	4%
White	<1%	91%	4%	76%
Two or more races	0%	3%	0%	3%

Note: ¹School-Wide Title I Program. Dashes indicate the information was unavailable.

Participant Recruitment and Selection. Research sites were recruited for this study via email, telephone, or in person by Acadience Learning. Prior to site recruitment and selection for this study, IRB approval was secured followed by school and/or district approval to conduct the study. Each school site had a coordinator to serve as the primary contact with Acadience Learning and facilitate the study. Consent letters for the study were sent to all parents of students eligible for participation. All students in grades K–4 who were learning to read in English were eligible to participate. Students at each site were selected for participation using a stratified random sampling approach at each grade such that the sample included about 50% At or Above Benchmark on their Acadience Reading beginning-of-year benchmark assessment, and the remaining 50% of students were Below or Well Below Benchmark. All school personnel who served as assessors and the site coordinator were invited to complete the electronic consumer feedback questionnaire. Consent letters for school personnel were distributed by the site coordinator. Prior to CFOL testing, student assent was obtained.

Measures and Training. The measures included the Acadience Reading Diagnostic CFOL measures, Acadience Reading measures, and a consumer feedback questionnaire. These measures are described in detail in the *Acadience Reading Diagnostic CFOL Validation Study Technical Report* (Powell-Smith et al., 2015) and the *Acadience Reading K–6 Technical Manual* (Good et al., 2013). All participating sites were already users of Acadience Reading. As such, no additional training on Acadience Reading was conducted for this study. Training for CFOL was conducted via a webinar that was approximately two hours in length. Three sites received training for most assessors via a live webinar. A few remaining assessors at these three sites were trained via a recorded webinar. The fourth site accessed the recorded training for all assessors. During the training, assessors were trained in the administration and scoring of the measures and provided opportunities for practice with feedback. The amount of time needed for the assessment and the potential for fatigue or frustration due to not using discontinue rules were discussed. Assessors were instructed to be sensitive to student fatigue and general affect and to take a break if needed and continue in another testing session.

Data Collection. Acadience Reading benchmark assessment data were collected as was typically conducted by each site. Data collection for CFOL was conducted individually and occurred mostly during the fall, however, some sites continued to collect data past the winter holidays in order to complete the testing with selected students. Testing time was expected to range from 45–60 minutes total per student, but reports from assessors indicated some longer assessment times. Students were administered all CFOL tasks designated for their grade level. No discontinue rules were used so that a factor analysis and item-response analyses could be conducted on all items in each section. A second trained assessor observed and scored CFOL assessments for about one third of the students tested as a means of obtaining inter-rater reliability data.

Data Management, Entry, and Reliability. All CFOL test forms were returned to Acadience Learning and inspected for completeness. In some instances, it appeared that assessors chose to discontinue a task or did not complete all items in some sections. This anomaly occurred for 36 students in one site only (8 in both kindergarten and first grade, 9 in both second and third grade, and 2 in fourth grade). We noted that assessors discontinued most frequently for the following three CFOL tasks and grade levels: Sentence Repetition in grades K–2, Making Words in grades K–2, and Reading Fluency in grade 3.

Research staff at Acadience Learning entered all data into an electronic spreadsheet. Separate Acadience Learning research staff then re-entered 55% of the data in a secondary spreadsheet for reliability. The two sets of data were compared and checked for accuracy electronically using an automated comparison formula. Any disagreements in the files resulted in re-examining the student test form to determine the accurate score and reconcile any disagreements in the electronic files. After checking for reliability, Acadience Learning research staff checked 100% of the data for scoring and tallying errors using automated mathematical formulas electronically. Any discrepancies between the automated scores and the original scores resulted in rescoring the student packet and correcting the electronic version. Any data-entry errors (e.g., typos such as entering the wrong score for an item) were corrected before analysis.

Data management procedures were followed to assure that a corrected and clean data set was used for data analysis. Data were checked for complete records, including scores and student ID numbers. Any issues with duplicate IDs were resolved prior to data analysis. We examined all of the data by grade for the presence of outlier students, patterns of guessing, missing scores, and/or invalid scores. We defined outlier students as those students whose earned scores at the extreme high end or extreme low end of the scoring range compared with the rest of the students in the sample. Although a couple of outlier scores were noted, the scores were consistent with the students' performance on other measures and there was no significant difference when comparing data sets with and without the outliers. As such, no outliers were removed.

Results. Descriptive data are presented first for the CFOL Validation Study, followed by results related to the validity and the reliability of the measures. Finally, we present the results of analyzing data from the consumer satisfaction questionnaire.

Descriptive statistics. Descriptive statistics for each CFOL task and grade are reported in Table 12.2. The results from the multiple comparisons procedure (Tukey's Honest Significant Difference) that tested for differences in the grade-level mean scores for each task are presented in the column titled "Group." Within each task, grade levels with different letters in the "Group" column had significantly different mean scores. Each task was analyzed independently of other tasks. Therefore, a grade level in Group A for one task is not comparable to a grade level in Group A in a different task. For example, in task A1, kindergarten and first grade had significantly different mean scores; kindergarten and first-grade students had significantly different scores in tasks A2, B1, and B2 as well. Grade-level mean scores with the same letter for each task were not significantly different from each other. Most tasks exhibited higher scores as grade level increased. Significant differences across all grades were found in the sections involving Story Coherence (A1, A2), Listening Comprehension (B1, B2), Reading Sentences with Homographs (C3), Morpheme Compounding (E1), and Vocabulary/Word Knowledge: Definitions (F1).

Table 12.2
CFOL Descriptive Statistics by Section

Section	Task	Grade	<i>n</i>	<i>Mean</i>	<i>SD</i>	Maximum Possible	Group*
A: Story Coherence/Text Structure							
A1	Story Telling	K	98	15.69	4.80		A
		1	99	17.09	4.19	25	B
A2	What Happens Next	K	102	4.64	1.92		A
		1	99	6.03	1.83	10	B
B: Listening Comprehension							
B1	Retell	K	102	6.48	4.76		A
		1	98	10.32	4.79	20	B
B2	Question and Answer	K	98	9.47	3.90		A
		1	99	12.42	3.71	18	B
C: Reading Comprehension							
C1	Paragraph Retell	2	90	14.72	4.55	20	-
C2	Sentences with Homophones	2	93	9.48	1.72	12	-
C3	Sentences with Homographs	2	92	4.49	2.45		A
		3	96	5.69	2.74		B
		4	84	6.79	2.15	10	C
C4	Passage Retell	3	92	15.23	3.54		A
		4	84	15.54	4.05	20	A
D: Syntactic Knowledge/Grammar							
D1	Matching Sentences to Pictures	K	102	5.02	1.68	8	-
D2	Use of Plurals	K	101	5.25	1.79		A
		1	99	6.14	2.19		B
		2	92	7.43	1.99		C
		3	94	8.02	1.81	10	C
D3	Use of Past Tense	K	99	2.41	1.51		A
		1	99	3.20	1.90		B
		2	92	4.58	1.94		C
		3	94	5.24	1.70	8	C
D4	Sentence Anagrams	K	101	1.77	2.12		A
		1	99	4.04	2.13		B
		2	93	5.94	1.44		C
		3	96	6.26	1.18	7	C
D5	Sentence Repetition	K	101	2.55	2.40		A
		1	98	3.85	2.41		B
		2	93	4.98	2.27		C
		3	96	4.82	2.09		C
		4	84	5.74	2.12	10	C

Note: Data were collected during the 2013–2014 school year. Levels not connected by the same letter within each task are significantly different at the $p < .04$ level. Maximum Possible indicates the maximum possible score for the task/section.

*Group is specific to each CFOL task.

Table 12.2
CFOL Descriptive Statistics by Section, continued

Section	Task	Grade	<i>n</i>	<i>Mean</i>	<i>SD</i>	Maximum Possible	Group*
E: Morphological Awareness							
E1	Morpheme Compounding	K	99	1.91	1.46		A
		1	99	2.87	1.51	5	B
E2	Sentence Completion (Decomposition, Derivation, & Inflection)	K	100	2.89	1.26		A
		1	99	3.41	1.21		A
		2	92	4.33	1.44		B
		3	94	5.35	1.62		C
		4	84	5.58	1.30	8	C
E3	Making Words	1	99	3.48	2.38		A
		2	86	4.98	2.15		B
		3	94	6.99	2.65		C
		4	84	7.93	2.89	15	C
F: Vocabulary/Word Knowledge							
F1	Definitions	K	99	4.72	3.55		A
		1	98	7.36	3.82		B
		2	90	9.61	4.19		C
		3	94	11.73	5.08		D
		4	84	14.14	4.98	24	E
F2	Multiple Meanings	K	99	6.43	3.87		A
		1	97	9.95	3.94		B
		2	88	11.89	4.46		C
		3	94	12.95	5.58		C
		4	84	16.25	3.65	20	D
F3	Figurative Language (Idioms)	1	98	3.39	3.25		A
		2	90	5.72	3.50		B
		3	94	8.13	3.16		C
		4	84	8.35	2.91	12	C
G: Reading Fluency							
G	Expository Passage	2	82	17.01	3.90		A
		3	84	16.81	4.18		A
		4	81	17.67	3.64	21	A
G	Narrative Passage	2	83	16.33	3.71		A
		3	85	16.41	3.84		A
		4	81	16.79	3.70	21	A

Note: Data were collected during the 2013–2014 school year. Levels not connected by the same letter within each task are significantly different at the $p < .04$ level. Maximum Possible indicates the maximum possible score for the task/section.

*Group is specific to each CFOL task.

Acadience Reading descriptive statistics are found in Table 12.3. When comparing these mean Acadience Reading scores to the benchmark goals for each measure at each grade level, on average the sample performed at or above their grade-level benchmark goals. Exceptions were Oral Reading Fluency Accuracy in grades 2 through 4 and Oral Reading Fluency Words Correct in grade 4.

Table 12.3
Acadience Reading Descriptive Statistics by Measure and Grade

Measure	Grade	<i>n</i>	Mean	SD
Letter Naming Fluency	K	96	19.52	13.50
	1	88	43.63	13.73
First Sound Fluency	K	96	14.67	12.76
	1	88	45.24	13.62
Phoneme Segmentation Fluency	1	88	30.68	19.96
	2	91	67.12	32.52
NWF Whole Words Read	1	88	3.35	7.38
	2	90	17.99	12.83
ORF Words Correct	2	91	56.43	27.25
	3	95	82.74	35.84
	4	82	87.74	34.99
	4	82	87.74	34.99
ORF Accuracy	2	91	87.97	14.69
	3	95	93.07	8.94
	4	82	93.40	8.99
ORF Retell	2	78	20.28	13.23
	3	79	28.19	15.94
	4	73	30.34	15.99
Maze	3	56	8.18	6.02
	4	48	17.00	7.01
Reading Composite Score	K	96	34.19	22.72
	1	88	119.55	36.17
	2	90	166.16	75.69
	3	40	251.83	109.35
	4	39	304.10	107.36

Note: Acadience Reading measures were administered at the beginning of the 2013–2014 school year. NWF = Nonsense Word Fluency. ORF = Oral Reading Fluency.

Tables 12.4 through 12.8 further describe the sample by comparing mean scores on the CFOL tasks of those who met or exceeded the Reading Composite Score (RCS) benchmark goal to those who did not. The percent of students at or above the RCS benchmark goal was 59% in kindergarten, 52% in first grade, and 62% in each second, third, and fourth grades. A *p* value of .05 or less indicates a significant difference between the mean scores on CFOL tasks for the Above Benchmark group and Below Benchmark group. For most CFOL tasks at each grade level, students who were at/above the RCS benchmark goal performed better on the CFOL task. These data are consistent with the validity data presented later in this chapter. However, at each grade level there were some CFOL tasks for which there were not significant differences (e.g., Definitions (F1) in kindergarten, Story Telling (A1) in first grade, Use of Plurals (D2) in second grade, Figurative Language (F3) in third grade, and Sentence Repetition (D5) in fourth grade).

Table 12.4

CFOL and Acadience Reading Kindergarten Descriptive Statistics by Measure and Benchmark Status

Measure	Above RCS Benchmark (59%)			Below RCS Benchmark (41%)			Maximum Possible	<i>p</i>
	<i>n</i>	Mean	<i>SD</i>	<i>n</i>	Mean	<i>SD</i>		
A1. Story Telling	55	16.91	4.26	37	14.22	5.14	25	.01
A2. What Happens Next	57	5.07	1.96	39	4.13	1.79	10	.02
B1. Retell	57	7.40	4.90	39	5.44	4.45	20	.05
B2. Question and Answer	55	10.24	3.81	39	8.54	3.59	18	.03
C1. Paragraph Retell	-	-	-	-	-	-	20	-
C2. Sentences with Homophones	-	-	-	-	-	-	12	-
C3. Sentences with Homographs	-	-	-	-	-	-	10	-
C4. Passage Retell	-	-	-	-	-	-	20	-
D1. Matching Sentences to Pictures	57	5.33	1.65	39	4.54	1.68	8	.02
D2. Use of Plurals	57	5.65	1.76	39	4.59	1.73	10	.00
D3. Use of Past Tense	56	2.71	1.68	38	1.89	1.09	8	.01
D4. Sentence Anagrams	57	2.37	2.37	39	0.97	1.42	7	.00
D5. Sentence Repetition	57	3.14	2.52	39	1.82	2.05	10	.01
E1. Morpheme Compounding	56	2.23	1.54	38	1.50	1.27	5	.01
E2. Sentence Completion (Decomposition, Derivation, & Inflection)	56	3.05	1.07	38	2.63	1.51	8	.14
E3. Making Words	-	-	-	-	-	-	15	-
F1. Definitions	55	5.18	3.74	38	4.34	3.22	24	.25
F2. Multiple Meanings	55	7.49	4.07	38	5.18	2.95	20	.00
F3. Figurative Language (Idioms)	-	-	-	-	-	-	12	-
G. Expository Passage	-	-	-	-	-	-	21	-
G. Narrative Passage	-	-	-	-	-	-	21	-
Letter Naming Fluency	57	27.28	11.41	39	8.18	6.44		.00
First Sound Fluency	57	21.37	11.68	39	4.87	6.35		.00
Reading Composite Score	57	48.65	17.57	39	13.05	7.95		.00

Note: Data were collected during the 2013–2014 school year. Ninety-two of the 102 students with CFOL data had complete Acadience Reading data required to compute a Reading Composite Score. RCS = Reading Composite Score.

Table 12.5*CFOL and Acadience Reading First Grade Descriptive Statistics by Measure and Benchmark Status*

Measure	Above RCS Benchmark (52%)			Below RCS Benchmark (48%)			Maximum Possible	p
	n	Mean	SD	n	Mean	SD		
A1. Story Telling	45	17.29	4.38	42	17.67	3.95	25	.67
A2. What Happens Next	46	6.33	1.71	42	5.64	1.81	10	.04
B1. Retell	45	11.36	4.95	42	9.24	4.61	20	.04
B2. Question and Answer	45	12.98	3.37	42	11.64	4.08	18	.10
C1. Paragraph Retell	-	-	-	-	-	-	20	-
C2. Sentences with Homophones	-	-	-	-	-	-	12	-
C3. Sentences with Homographs	-	-	-	-	-	-	10	-
C4. Passage Retell	-	-	-	-	-	-	20	-
D1. Matching Sentences to Pictures	-	-	-	-	-	-	8	-
D2. Use of Plurals	45	6.58	2.28	42	5.69	2.15	10	.07
D3. Use of Past Tense	45	3.51	2.01	42	2.76	1.65	8	.06
D4. Sentence Anagrams	45	4.76	2.08	42	3.43	1.84	7	.00
D5. Sentence Repetition	45	4.24	2.28	41	3.32	2.56	10	.08
E1. Morpheme Compounding	45	3.09	1.43	42	2.52	1.50	5	.08
E2. Sentence Completion (Decomposition, Derivation, & Inflection)	45	3.67	1.33	42	3.14	1.03	8	.04
E3. Making Words	45	3.96	2.15	42	3.02	2.71	15	.08
F1. Definitions	44	7.89	4.27	42	6.45	3.00	24	.07
F2. Multiple Meanings	43	10.53	3.48	42	9.00	4.37	20	.08
F3. Figurative Language (Idioms)	44	4.05	3.48	42	2.71	2.73	12	.05
G. Expository Passage	-	-	-	-	-	-	21	-
G. Narrative Passage	-	-	-	-	-	-	21	-
Letter Naming Fluency	46	52.67	10.19	42	33.71	9.69		.00
Phoneme Segmentation Fluency	46	51.46	9.93	42	38.43	13.95		.00
NWF Correct Letter Sounds	46	40.83	22.54	42	19.57	6.80		.00
NWF Whole Words Read	46	5.65	9.52	42	0.83	1.95		.00
Reading Composite Score	46	144.96	29.76	42	91.71	16.94		.00

Note: Maximum Possible indicates the maximum possible score for the task/section. Eighty-eight of the 100 students with CFOL data had complete Acadience Reading data required to compute a Reading Composite Score. RCS = Reading Composite Score. NWF = Nonsense Word Fluency.

Table 12.6*CFOL and Acadience Reading Second Grade Descriptive Statistics by Measure and Benchmark Status*

Measure	Above RCS Benchmark (62%)			Below RCS Benchmark (38%)			Maximum Possible	p
	n	Mean	SD	n	Mean	SD		
A1. Story Telling	-	-	-	-	-	-	25	-
A2. What Happens Next	-	-	-	-	-	-	10	-
B1. Retell	-	-	-	-	-	-	20	-
B2. Question and Answer	-	-	-	-	-	-	18	-
C1. Paragraph Retell	55	15.75	3.64	32	12.84	5.51	20	.01
C2. Sentences with Homophones	56	10.07	1.52	34	8.38	1.52	12	.00
C3. Sentences with Homographs	56	5.41	2.18	33	2.73	1.91	10	.00
C4. Passage Retell	-	-	-	-	-	-	20	-
D1. Matching Sentences to Pictures	-	-	-	-	-	-	8	-
D2. Use of Plurals	56	7.59	2.06	33	7.03	1.88	10	.20
D3. Use of Past Tense	56	5.00	1.81	33	3.73	1.94	8	.00
D4. Sentence Anagrams	56	6.30	1.16	34	5.35	1.61	7	.00
D5. Sentence Repetition	56	5.27	2.11	34	4.47	2.55	10	.13
E1. Morpheme Compounding	-	-	-	-	-	-	5	-
E2. Sentence Completion (Decomposition, Derivation, & Inflection)	56	4.52	1.50	33	3.82	1.10	8	.01
E3. Making Words	54	5.33	2.26	29	4.14	1.71	15	.01
F1. Definitions	55	10.07	4.07	32	8.84	4.42	24	.20
F2. Multiple Meanings	55	12.82	4.17	30	9.93	4.55	20	.01
F3. Figurative Language (Idioms)	55	5.98	3.18	32	4.91	3.86	12	.19
G. Expository Passage	54	17.33	3.05	26	13.92	3.93	21	.00
G. Narrative Passage	54	18.13	3.07	25	14.40	4.41	21	.00
NWF Correct Letter Sounds	56	82.71	29.52	34	39.97	13.58		.00
NWF Whole Words Read	56	24.61	11.66	34	7.09	4.56		.00
ORF Words Correct	56	71.71	18.50	34	29.44	13.90		.00
ORF Accuracy	56	95.57	2.80	34	75.09	17.38		.00
ORF Retell	52	25.13	12.45	25	9.64	7.44		.00
Reading Composite Score	56	214.25	43.37	34	86.94	43.88		.00

Note: Data were collected during the 2013–2014 school year. Ninety of the 96 students with CFOL data had complete Acadience Reading data required to compute a Reading Composite Score. RCS = Reading Composite Score. NWF = Nonsense Word Fluency. ORF = Oral Reading Fluency. Dashes indicate the measure is not administered at the specified grade level.

Table 12.7*CFOL and Acadience Reading Third Grade Descriptive Statistics by Measure and Benchmark Status*

Measure	Above ORF Words Correct Benchmark (62%)			Below ORF Words Correct Benchmark (38%)			Maximum Possible	p
	n	Mean	SD	n	Mean	SD		
A1. Story Telling	-	-	-	-	-	-	25	-
A2. What Happens Next	-	-	-	-	-	-	10	-
B1. Retell	-	-	-	-	-	-	20	-
B2. Question and Answer	-	-	-	-	-	-	18	-
C1. Paragraph Retell	-	-	-	32	12.84	5.51	20	.01
C2. Sentences with Homophones	-	-	-	34	8.38	1.52	12	.00
C3. Sentences with Homographs	59	6.56	2.51	33	2.73	1.91	10	.00
C4. Passage Retell	59	16.14	3.26	-	-	-	20	-
D1. Matching Sentences to Pictures	-	-	-	-	-	-	8	-
D2. Use of Plurals	58	8.26	1.47	33	7.03	1.88	10	.20
D3. Use of Past Tense	58	5.53	1.68	33	3.73	1.94	8	.00
D4. Sentence Anagrams	59	6.51	0.80	34	5.35	1.61	7	.00
D5. Sentence Repetition	59	5.08	1.90	34	4.47	2.55	10	.13
E1. Morpheme Compounding	-	-	-	-	-	-	5	-
E2. Sentence Completion (Decomposition, Derivation, & Inflection)	58	5.69	1.37	33	3.82	1.10	8	.01
E3. Making Words	58	7.24	2.86	29	4.14	1.71	15	.01
F1. Definitions	58	12.47	4.97	32	8.84	4.42	24	.20
F2. Multiple Meanings	58	13.66	5.30	30	9.93	4.55	20	.01
F3. Figurative Language (Idioms)	58	8.34	2.79	32	4.91	3.86	12	.19
G. Expository Passage	53	17.57	3.38	31	14.61	3.84	21	.00
G. Narrative Passage	52	18.29	3.17	32	14.41	4.54	21	.00
ORF Words Correct	59	103.64	27.92	36	48.47	14.59		.00
ORF Accuracy	59	96.69	2.15	36	87.14	12.19		.00
ORF Retell	50	31.46	17.51	29	22.55	10.91		.02
Maze	36	10.81	5.79	20	3.45	2.56		.00
Reading Composite Score	27	300.33	89.57	13	151.08	72.03		.00

Note: Data were collected during the 2013–2014 school year. Ninety-five of the 96 students with CFOL data had ORF data. ORF = Oral Reading Fluency. Dashes indicate the measure is not administered at the specified grade level.

Table 12.8*CFOL and Acadience Reading Fourth Grade Descriptive Statistics by Measure and Benchmark Status*

Measure	Above ORF Words Correct Benchmark (44%)			Below ORF Words Correct Benchmark (56%)			Maximum Possible	p
	n	Mean	SD	n	Mean	SD		
A1. Story Telling	-	-	-	-	-	-	25	-
A2. What Happens Next	-	-	-	-	-	-	10	-
B1. Retell	-	-	-	-	-	-	20	-
B2. Question and Answer	-	-	-	-	-	-	18	-
C1. Paragraph Retell	-	-	-	-	-	-	20	-
C2. Sentences with Homophones	-	-	-	-	-	-	12	-
C3. Sentences with Homographs	36	7.69	1.72	46	6.09	2.22	10	.00
C4. Passage Retell	36	16.67	3.01	46	14.63	4.61	20	.02
D1. Matching Sentences to Pictures	-	-	-	-	-	-	8	-
D2. Use of Plurals	-	-	-	-	-	-	10	-
D3. Use of Past Tense	-	-	-	-	-	-	8	-
D4. Sentence Anagrams	-	-	-	-	-	-	7	-
D5. Sentence Repetition	36	5.92	2.39	46	5.59	1.94	10	.50
E1. Morpheme Compounding	-	-	-	-	-	-	5	-
E2. Sentence Completion (Decomposition, Derivation, & Inflection)	36	5.86	1.10	46	5.39	1.42	8	.10
E3. Making Words	36	8.67	2.95	46	7.39	2.77	15	.05
F1. Definitions	36	15.67	4.58	46	12.89	5.03	24	.01
F2. Multiple Meanings	36	17.11	3.08	46	15.54	4.00	20	.05
F3. Figurative Language (Idioms)	36	8.89	2.88	46	7.87	2.92	12	.12
G. Expository Passage	35	18.83	2.58	44	15.27	3.74	21	.00
G. Narrative Passage	35	19.69	1.86	44	16.07	3.98	21	.00
ORF Words Correct	36	118.58	24.77	46	63.61	19.24		.00
ORF Accuracy	36	97.36	1.87	46	90.30	10.98		.00
ORF Retell	33	40.30	15.00	40	22.13	11.59		.00
Maze	24	20.96	7.29	24	13.04	3.83		.00
Reading Composite Score	21	378.52	65.65	18	217.28	76.61		.00

Note: Data were collected during the 2013–2014 school year. Eighty-two of the 84 students with CFOL data had ORF data. ORF = Oral Reading Fluency. Dashes indicate the measure is not administered at the specified grade level.

Validity. Several types of validity data are provided in this section of the chapter. First, we provide information on criterion-related validity in the form of correlations with Acadience Reading measures. Next, validity is addressed in three distinct ways that relate to the construction of the measures: (a) correlations among the CFOL sections, (b) item-response analyses, and (c) factor analysis.

Criterion-related validity (either concurrent or predictive) is the degree to which performance on a criterion measure can be estimated from performance on an assessment. Validity is measured as a correlation between the criterion and the assessment. The guidelines from Hopkins (2006), shown in Table 12.9, are used when describing the results.

Table 12.9
Validity Estimate Descriptors

Validity Correlation Range	Descriptor
.70 and above	Strong
.50 – .69	Moderate-Strong
.30 – .49	Moderate
.10 – .29	Small
.09 or less	Very Small

Correlations with Acadience Reading. Tables 12.10 through 12.13 report concurrent and predictive validity correlations between each CFOL task and the Acadience Reading measures at each grade. Acadience Reading measures were administered at the beginning and middle of the academic year, while the CFOL measures were administered near the beginning of the year in most instances. Where significant, correlations for Section A tasks range from small (.20 with Letter Naming Fluency (LNF)) to moderate (.32 with First Sound Fluency (FSF)) in kindergarten. Tasks B1 and B2 correlations range from small to moderate with LNF and the Reading Composite Score (RCS), respectively. Correlations with tasks in Section C range from .21 to .70, with the strongest correlations between the RCS in second and third grade and Sentences with Homographs (C3).

Correlations for tasks in Section D range from small to moderate in kindergarten through second grade, and small to moderate-strong in third and fourth grade (.21 to .36 in kindergarten, .24 to .46 in first grade, .21 to .49 in second, .22 to .70 in third, and .31 to .60 in fourth). In Section E, Morpheme Compounding (E1) was most strongly correlated with FSF (.46 in kindergarten), Sentence Completion (E2) was most strongly correlated with Oral Reading Fluency (ORF) and the RCS (.49, .55, respectively, in third grade), and Making Words (E3) was most strongly correlated with ORF Accuracy (.49 in first, .50 in fourth grade). Vocabulary and Word Knowledge (Section F) correlated most strongly with Maze in third and fourth grades (.63, .59). Finally, Reading Fluency (Section G) was moderately to strongly correlated with ORF and the RCS (.38 to .79), with the strongest correlations occurring in the middle of the year in second grade.

Table 12.10*Correlations Between CFOL Tasks in Sections A and B and Acadience Reading Measures*

Grade	Acadience Reading Measure— Time of Year	CFOL Task			
		A1	A2	B1	B2
K	LNF—beginning	.20	.20	.27	.21
	FSF—beginning	.32	.24	.30	.32
	RCS—beginning	.30	.26	.33	.30
	LNF—middle	.12	.18 [†]	.21	.26
	FSF—middle	.23	.17 [†]	.17 [†]	.30
	PSF—middle	.18 [†]	.22	.32	.38
	NWF CLS—middle	.08 [†]	.04 [†]	.12 [†]	.15 [†]
	NWF WWR—middle	-.07 [†]	-.09 [†]	-.09 [†]	-.06 [†]
	RCS—middle	.19 [†]	.19 [†]	.26	.34
First	LNF—beginning	-.02 [†]	.14 [†]	.19 [†]	.16 [†]
	PSF—beginning	-.03 [†]	.19 [†]	.21 [†]	.15 [†]
	NWF CLS—beginning	.01 [†]	.08 [†]	.12 [†]	.14 [†]
	NWF WWR—beginning	.00 [†]	.10 [†]	.14 [†]	.16 [†]
	RCS—beginning	-.01 [†]	.17 [†]	.22	.20 [†]
	NWF CLS—middle	.06 [†]	.12 [†]	.19 [†]	.18 [†]
	NWF WWR—middle	-.06 [†]	.10 [†]	.17 [†]	.16 [†]
	ORF—middle	.10 [†]	.20 [†]	.32	.23
	ACC—middle	.13 [†]	.26	.34	.28
	Retell—middle	.18 [†]	.08 [†]	.29	.28
RCS—middle	.08 [†]	.19 [†]	.30	.26	

Note: In most instances, CFOL was administered near the beginning of year. Unless otherwise marked, correlation significant at $p < .05$; [†] = not significant. LNF = Letter Naming Fluency. FSF = First Sound Fluency. RCS = Reading Composite Score. PSF = Phoneme Segmentation Fluency. NWF CLS = Nonsense Word Fluency Correct Letter Sounds. NWF WWR = Nonsense Word Fluency Whole Words Read. ORF = Oral Reading Fluency. ACC = Accuracy.

Table 12.11*Correlations Between CFOL Tasks in Section C and Acadience Reading Measures*

Grade	Acadience Reading Measure— Time of Year	CFOL Task			
		C1	C2	C3	C4
Second	NWF CLS—beginning	.31	.52	.51	-
	NWF WWR—beginning	.30	.52	.52	-
	ORF—beginning	.38	.62	.56	-
	ACC—beginning	.51	.46	.53	-
	Retell—beginning	.27	.25	.36	-
	RCS—beginning	.44	.62	.61	-
	ORF—middle	.52	.60	.65	-
	ACC—middle	.63	.36	.47	-
	Retell—middle	.49	.32	.39	-
	RCS—middle	.61	.50	.70	-
Third	ORF—beginning	-	-	.59	.36
	ACC—beginning	-	-	.55	.34
	Retell—beginning	-	-	.52	.36
	Maze—beginning	-	-	.57	.19 [†]
	RCS—beginning	-	-	.69	.33
	ORF—middle	-	-	.63	.40
	ACC—middle	-	-	.58	.34
	Retell—middle	-	-	.27	.30
	Maze—middle	-	-	.65	.19 [†]
	RCS—middle	-	-	.61	.32 [†]
Fourth	ORF—beginning	-	-	.43	.44
	ACC—beginning	-	-	.58	.65
	Retell—beginning	-	-	.21	.36
	Maze—beginning	-	-	.35	.28
	RCS—beginning	-	-	.46	.47
	ORF—middle	-	-	.46	.37
	ACC—middle	-	-	.49	.46
	Retell—middle	-	-	.36	.38
	Maze—middle	-	-	.40	.46
	RCS—middle	-	-	.59	.51

Note: In most instances, CFOL was administered near the beginning of year. Unless otherwise marked, correlation significant at $p < .05$;

[†] = not significant. RCS = Reading Composite Score. NWF CLS = Nonsense Word Fluency Correct Letter Sounds. NWF WWR = Nonsense Word Fluency Whole Words Read. ORF = Oral Reading Fluency. ACC = Accuracy. Dashes indicate the CFOL task is not administered at the specified grade level.

Table 12.12*Correlations Between CFOL Tasks in Section D and Acadience Reading Measures*

Grade	Acadience Reading Measure— Time of Year	CFOL Task				
		D1	D2	D3	D4	D5
K	LNF—beginning	.24	.15 [†]	.25	.30	.23
	FSF—beginning	.23	.29	.35	.32	.38
	RCS—beginning	.27	.25	.34	.36	.35
	LNF—middle	.31	.15 [†]	.22	.30	.27
	FSF—middle	.33	.19 [†]	.21	.27	.20 [†]
	PSF—middle	.26	.32	.26	.29	.23
	NWF CLS—middle	.24 [†]	.13 [†]	.13 [†]	.26	.27
	NWF WWR—middle	-.04 [†]	.02 [†]	-.07 [†]	.09 [†]	.11 [†]
	RCS—middle	.34	.25	.25	.34	.29
First	LNF—beginning	-	.02 [†]	.03 [†]	.20 [†]	.10 [†]
	PSF—beginning	-	.34	.28	.35	.37
	NWF CLS—beginning	-	.09 [†]	.07 [†]	.28	.13 [†]
	NWF WWR—beginning	-	.14 [†]	.10 [†]	.24	.09 [†]
	RCS—beginning	-	.18 [†]	.15 [†]	.36	.25
	NWF CLS—middle	-	.21 [†]	.20 [†]	.26	.30
	NWF WWR—middle	-	.28	.21 [†]	.25	.25
	ORF—middle	-	.20 [†]	.09 [†]	.42	.26
	ACC—middle	-	.29	.14 [†]	.46	.32
	Retell—middle	-	.30	.16 [†]	.41	.17 [†]
	RCS—middle	-	.27	.18 [†]	.40	.32
	Second	NWF CLS—beginning	-	.32	.37	.39
NWF WWR—beginning		-	.25	.40	.36	.06 [†]
ORF—beginning		-	.32	.38	.41	.21
ACC—beginning		-	.32	.41	.42	.20 [†]
Retell—beginning		-	.20 [†]	.19 [†]	.27	.16 [†]
RCS—beginning		-	.32	.44	.44	.18 [†]
ORF—middle		-	.38	.39	.49	.15 [†]
ACC—middle		-	.34	.50	.45	.01 [†]
Retell—middle		-	.12 [†]	.12 [†]	.29	.23 [†]
RCS—middle	-	.32	.43	.58	.08 [†]	

Note: In most instances, CFOL was administered near the beginning of year. Unless otherwise marked, correlation significant at $p < .05$; [†] = not significant.

LNF = Letter Naming Fluency. FSF = First Sound Fluency. RCS = Reading Composite Score. PSF = Phoneme Segmentation Fluency. NWF CLS = Nonsense Word Fluency Correct Letter Sounds. NWF WWR = Nonsense Word Fluency Whole Words Read. ORF = Oral Reading Fluency. ACC = Accuracy. Dashes indicate the CFOL task is not administered at the specified grade level.

Table 12.12*Correlations Between CFOL Tasks in Section D and Acadience Reading Measures, continued*

Grade	Acadience Reading Measure— Time of Year	CFOL Task				
		D1	D2	D3	D4	D5
Third	ORF—beginning	-	.28	.37	.40	.35
	ACC—beginning	-	.50	.43	.67	.42
	Retell—beginning	-	.38	.22	.38	.40
	Maze—beginning	-	.21 [†]	.44	.24 [†]	.30
	RCS—beginning	-	.36	.29 [†]	.39	.21 [†]
	ORF—middle	-	.32	.41	.43	.41
	ACC—middle	-	.59	.47	.70	.44
	Retell—middle	-	.08 [†]	-.04 [†]	.15 [†]	.39
	Maze—middle	-	.40	.43	.43	.33
	RCS—middle	-	.08 [†]	.14 [†]	.13 [†]	.20 [†]
Fourth	ORF—beginning	-	.28	.37	.40	.35
	ACC—beginning	-	.50	.43	.67	.42
	Retell—beginning	-	.38	.22	.38	.40
	Maze—beginning	-	.21 [†]	.44	.24 [†]	.30
	RCS—beginning	-	.36	.29 [†]	.39	.21 [†]
	ORF—middle	-	.32	.41	.43	.41
	ACC—middle	-	.59	.47	.70	.44
	Retell—middle	-	.08 [†]	-.04 [†]	.15 [†]	.39
	Maze—middle	-	.40	.43	.43	.33
	RCS—middle	-	.08 [†]	.14 [†]	.13 [†]	.20 [†]

Note: In most instances, CFOL was administered near the beginning of year. Unless otherwise marked, correlation significant at $p < .05$; [†] = not significant. LNF = Letter Naming Fluency. FSF = First Sound Fluency. RCS = Reading Composite Score. PSF = Phoneme Segmentation Fluency. NWF CLS = Nonsense Word Fluency Correct Letter Sounds. NWF WWR = Nonsense Word Fluency Whole Words Read. ORF = Oral Reading Fluency. ACC = Accuracy. Dashes indicate the CFOL task is not administered at the specified grade level.

Table 12.13*Correlations Between CFOL Tasks in Sections E Through G and Acadience Reading Measures*

Grade	Acadience Reading Measure— Time of Year	CFOL Task							
		E1	E2	E3	F1	F2	F3	Ge	Gn
K	LNF—beginning	.31	.13 [†]	-	.13 [†]	.34	-	-	-
	FSF—beginning	.46	.26	-	.26	.40	-	-	-
	RCS—beginning	.44	.23	-	.22	.42	-	-	-
	LNF—middle	.34	.15 [†]	-	-.01 [†]	.14 [†]	-	-	-
	FSF—middle	.32	.20 [†]	-	.08 [†]	.15 [†]	-	-	-
	PSF—middle	.32	.22	-	.14 [†]	.10 [†]	-	-	-
	NWF CLS—middle	.39	.13 [†]	-	.03 [†]	.08 [†]	-	-	-
	NWF WWR—middle	.15 [†]	-.05 [†]	-	.03 [†]	-.15 [†]	-	-	-
	RCS—middle	.42	.22	-	.08 [†]	.14 [†]	-	-	-
First	LNF—beginning	.01 [†]	.03 [†]	.15 [†]	.22	.15 [†]	.17 [†]	-	-
	PSF—beginning	.26	.28	.18 [†]	.33	.36	.30	-	-
	NWF CLS—beginning	.08 [†]	.14 [†]	.16 [†]	.32	.24	.16 [†]	-	-
	NWF WWR—beginning	.19 [†]	.24	.24	.23	.15 [†]	.15 [†]	-	-
	RCS—beginning	.15 [†]	.19 [†]	.21	.38	.32	.27	-	-
	NWF CLS—middle	.12 [†]	.19 [†]	.24	.36	.28	.21 [†]	-	-
	NWF WWR—middle	.19 [†]	.25	.24	.39	.28	.28	-	-
	ORF—middle	.14 [†]	.15 [†]	.36	.32	.22	.28	-	-
	ACC—middle	.29	.23	.49	.32	.36	.33	-	-
Second	Retell—middle	.25	.26	.19 [†]	.42	.27	.38	-	-
	RCS—middle	.21 [†]	.22	.39	.39	.32	.30	-	-
	NWF CLS—beginning	-	.30	.13 [†]	.22	.27	.27	.51	.52
	NWF WWR—beginning	-	.29	.13 [†]	.16 [†]	.21	.24	.48	.47
	ORF—beginning	-	.39	.23	.20 [†]	.29	.24	.60	.59
	ACC—beginning	-	.30	.30	.14 [†]	.37	.22	.58	.60
	Retell—beginning	-	.25	.30	.23	.33	.15 [†]	.49	.43
	RCS—beginning	-	.35	.24	.20 [†]	.33	.25	.61	.61
	ORF—middle	-	.42	.27	.13 [†]	.32	.24	.63	.60
ACC—middle	-	.32	.34	.21 [†]	.45	.14 [†]	.72	.66	
Retell—middle	-	.32	.41	.16 [†]	.25 [†]	-.03 [†]	.46	.38	
RCS—middle	-	.47	.42	.18 [†]	.34	.13 [†]	.79	.72	

Note: In most instances, CFOL was administered near the beginning of year. Unless otherwise marked, correlation significant at $p < .05$; [†] = not significant.

Ge = Expository Passage. Gn = Narrative Passage. LNF = Letter Naming Fluency. FSF = First Sound Fluency. RCS = Reading Composite Score.

PSF = Phoneme Segmentation Fluency. NWF CLS = Nonsense Word Fluency Correct Letter Sounds. NWF WWR = Nonsense Word Fluency Whole Words Read. ORF = Oral Reading Fluency. ACC = Accuracy. Dashes indicate the CFOL task is not administered at the specified grade level.

Table 12.13

Correlations Between CFOL Tasks in Sections E Through G and Acadience Reading Measures, continued

Grade	Acadience Reading Measure— Time of Year	CFOL Task							
		E1	E2	E3	F1	F2	F3	Ge	Gn
Third	ORF—beginning	-	.42	.19 [†]	.31	.31	.24	.45	.48
	ACC—beginning	-	.38	.22	.30	.35	.40	.28	.36
	Retell—beginning	-	.46	.35	.39	.47	.42	.41	.36
	Maze—beginning	-	.38	.12 [†]	.20 [†]	.28	.20 [†]	.22 [†]	.25 [†]
	RCS—beginning	-	.46	.30 [†]	.29 [†]	.44	.39	.38	.35
	ORF—middle	-	.49	.28	.43	.40	.25	.60	.60
	ACC—middle	-	.46	.23	.35	.53	.43	.54	.56
	Retell—middle	-	.34	.15 [†]	.39	.37	.24 [†]	.46	.46
	Maze—middle	-	.43	-.10	.20 [†]	.63	.28 [†]	.55	.59
	RCS—middle	-	.55	.10 [†]	.09 [†]	.40 [†]	.27 [†]	.76	.66
Fourth	ORF—beginning	-	.22	.32	.34	.27	.21 [†]	.50	.50
	ACC—beginning	-	.12 [†]	.36	.29	.23	.15 [†]	.59	.60
	Retell—beginning	-	.06 [†]	.19 [†]	.27	.27	.34	.37	.41
	Maze—beginning	-	.15 [†]	.19 [†]	.30	.49	.43	.56	.58
	RCS—beginning	-	.17 [†]	.28 [†]	.42	.42	.43	.57	.61
	ORF—middle	-	.25 [†]	.35	.25 [†]	.35	.24 [†]	.56	.54
	ACC—middle	-	.28	.50	.37	.36	.26	.64	.66
	Retell—middle	-	.10 [†]	.08 [†]	.32	.37	.23 [†]	.37	.35
	Maze—middle	-	.37 [†]	.34 [†]	.15 [†]	.59	.34 [†]	.60	.62
	RCS—middle	-	.36 [†]	.32 [†]	.36 [†]	.52	.33 [†]	.65	.68

Note: In most instances, CFOL was administered near the beginning of year. Unless otherwise marked, correlation significant at $p < .05$; [†] = not significant.

Ge = Expository Passage. Gn = Narrative Passage. LNF = Letter Naming Fluency. FSF = First Sound Fluency. RCS = Reading Composite Score.

PSF = Phoneme Segmentation Fluency. NWF CLS = Nonsense Word Fluency Correct Letter Sounds. NWF WWR = Nonsense Word Fluency Whole Words Read. ORF = Oral Reading Fluency. ACC = Accuracy. Dashes indicate the CFOL task is not administered at the specified grade level.

Correlations between CFOL sections. Correlations between the CFOL sections at each grade level are presented in Tables 12.14 through 12.18. Correlations between the Reading Comprehension (C) tasks and the Reading Fluency tasks (G) were moderate to moderate-strong (.43 to .63) for all grades tested on both of these tasks. In kindergarten, correlations were strongest between the listening comprehension measures (B1 and B2; .66) and Sentence Anagrams (D4) and Morpheme Compounding (E1) (.65). Other correlational relationships ranged from small to moderate-strong. First-grade correlations were strongest between tasks within Sections D and F, ranging from .55 to .66 in Section F: Vocabulary/Word Knowledge, and .27 to .64 in Section D: Syntactical Knowledge/Grammar. Within section correlations in second grade were mostly moderate-strong or better in Sections C: Reading Comprehension, F: Vocabulary/Word Knowledge, and G: Reading Fluency. In addition, correlations between Reading Comprehension tasks (C) and Reading Fluency tasks (G) ranged from .49 to .58. Third-grade correlations between Multiple Meanings (F2) and Figurative Language (F3) were moderate to moderately-strong related to Sentences with Homographs (C3),

Passage Retell (C4), and Use of Plurals (D2) (.45–.58). The strongest correlational relationships in fourth grade were between tasks of the same section (F, G) and Sections C: Reading Comprehension and G: Reading Fluency (.52–.63).

Table 12.14
CFOL Kindergarten Correlations

Section	A1	A2	B1	B2	D1	D2	D3	D4	D5	E1	E2	F1
A2	.33	-										
B1	.23	.42	-									
B2	.29	.50	.66	-								
D1	.15 [†]	.38	.37	.53	-							
D2	.26	.35	.31	.48	.36	-						
D3	.29	.45	.55	.49	.47	.45	-					
D4	.31	.44	.54	.59	.48	.52	.50	-				
D5	.24	.37	.42	.46	.31	.37	.28	.60	-			
E1	.22	.42	.47	.50	.49	.45	.35	.65	.45	-		
E2	.11 [†]	.24	.39	.42	.28	.44	.36	.46	.28	.41	-	
F1	.26	.40	.48	.42	.17 [†]	.29	.39	.45	.43	.36	.46	-
F2	.30	.47	.61	.49	.38	.35	.52	.48	.41	.46	.45	.55

Note: Data were collected during the 2013–2014 school year. Sample sizes for pairwise correlations between CFOL measures range from 94 to 102. Unless otherwise marked, correlation significant at $p < .05$; [†] = not significant.

Table 12.15
CFOL First-Grade Correlations

Section	A1	A2	B1	B2	D2	D3	D4	D5	E1	E2	E3	F1	F2
A2	.22	-											
B1	.19 [†]	.41	-										
B2	.17 [†]	.42	.52	-									
D2	.09 [†]	.14 [†]	.40	.50	-								
D3	-.03 [†]	.20 [†]	.41	.49	.64	-							
D4	.28	.14 [†]	.48	.33	.45	.27	-						
D5	-.01 [†]	.27	.53	.48	.34	.44	.36	-					
E1	-.09 [†]	.31	.47	.48	.57	.55	.34	.28	-				
E2	.04 [†]	.16 [†]	.38	.35	.54	.55	.34	.42	.52	-			
E3	.19 [†]	.31	.38	.43	.36	.29	.46	.29	.46	.33	-		
F1	.10 [†]	.14 [†]	.41	.37	.39	.43	.42	.41	.39	.45	.30	-	
F2	.03 [†]	.23	.50	.49	.52	.49	.45	.50	.44	.46	.30	.64	-
F3	.09 [†]	.29	.49	.42	.53	.44	.34	.29	.46	.49	.22	.66	.55

Note: Data were collected during the 2013–2014 school year. Sample sizes for pairwise correlations between CFOL measures range from 93 to 99. Unless otherwise marked, correlation significant at $p < .05$; [†] = not significant.

Table 12.16
CFOL Second-Grade Correlations

Section	C1	C2	C3	D2	D3	D4	D5	E2	E3	F1	F2	F3	Ge
C2	.53	-											
C3	.59	.51	-										
D2	.43	.50	.36	-									
D3	.52	.52	.60	.55	-								
D4	.54	.41	.52	.51	.45	-							
D5	.42	.28	.41	.21	.26	.33	-						
E2	.39	.45	.55	.59	.53	.48	.29	-					
E3	.37	.21 [†]	.30	.11 [†]	.37	.21 [†]	.05 [†]	.29	-				
F1	.31	.18 [†]	.15 [†]	.49	.31	.32	.00 [†]	.34	.13 [†]	-			
F2	.36	.35	.32	.61	.51	.44	.06 [†]	.47	.35	.56	-		
F3	.32	.31	.35	.44	.42	.25	.03 [†]	.45	.28	.50	.53	-	
Ge	.58	.51	.56	.47	.38	.54	.35	.51	.17 [†]	.40	.42	.36	-
Gn	.55	.49	.57	.46	.35	.51	.36	.42	.15 [†]	.37	.39	.36	.92

Note: Data were collected during the 2013–2014 school year. Sample sizes for pairwise correlations between CFOL measures range from 77 to 93. Unless otherwise marked, correlation significant at $p < .05$; [†] = not significant. Ge = Section G Expository. Gn = Section G Narrative.

Table 12.17
CFOL Third-Grade Correlations

Section	C3	C4	D2	D3	D4	D5	E2	E3	F1	F2	F3	Ge
C4	.41	-										
D2	.48	.28	-									
D3	.54	.33	.48	-								
D4	.58	.35	.45	.37	-							
D5	.48	.38	.44	.43	.42	-						
E2	.49	.46	.52	.53	.40	.42	-					
E3	.18 [†]	.27	.12 [†]	.10 [†]	.20 [†]	.17 [†]	.21	-				
F1	.44	.31	.42	.26	.38	.38	.45	.17 [†]	-			
F2	.58	.51	.52	.39	.49	.38	.51	.10 [†]	.52	-		
F3	.54	.45	.53	.42	.55	.40	.50	.14 [†]	.51	.63	-	
Ge	.56	.43	.23	.32	.20 [†]	.22	.31	.17 [†]	.25	.45	.30	-
Gn	.49	.50	.25	.32	.25	.32	.39	.25	.23	.37	.25	.85

Note: Data were collected during the 2013–2014 school year. Sample sizes for pairwise correlations between CFOL measures range from 82 to 96. Unless otherwise marked, correlation significant at $p < .05$; [†] = not significant. Ge = Section G Expository. Gn = Section G Narrative.

Table 12.18
CFOL Fourth-Grade Correlations

Section	C3	C4	D5	E2	E3	F1	F2	F3	Ge
C4	.44	-							
D5	.14 [†]	.27	-						
E2	.30	.29	.17 [†]	-					
E3	.25	.39	.14 [†]	.01 [†]	-				
F1	.34	.36	.30	.27	.26	-			
F2	.32	.44	.46	.24	.30	.46	-		
F3	.19 [†]	.32	.44	.17 [†]	.21 [†]	.52	.59	-	
Ge	.63	.57	.31	.32	.27	.40	.47	.38	-
Gn	.61	.52	.28	.33	.21 [†]	.35	.48	.40	.91

Note: Data were collected during the 2013–2014 school year. Sample sizes for pairwise correlations between CFOL measures range from 81 to 84. Unless otherwise marked, correlation significant at $p < .05$; [†] = not significant.
Ge = Section G Expository. Gn = Section G Narrative.

Validity of the internal structure of CFOL. A number of changes were made to Acadience Reading Diagnostic CFOL as a result of the CFOL Pilot study, the Pilot-2 study, and the CFOL Validation study. These changes were made as a result of research, theory, and analysis with the purpose of improving upon the instrument’s ability to measure the varying aspects of a single latent construct: reading comprehension. A key aspect of validating these changes (which are described later in this chapter) is investigating the internal structure of CFOL, and verifying that the relationships among test items and task sections correspond to a single latent construct (Bachman & Palmer, 1996). To verify the internal structure of CFOL, we evaluated estimates of ability level and difficulty from the item response analysis and evaluated the factor structure from the confirmatory factor analysis.

Item response analysis. Items were evaluated through various statistics: raw score, level of response, difficulty parameters, discrimination parameters, information, the area under the information curve, average ability level, and Cronbach’s alpha. Not every metric was used to evaluate each section. Descriptions of these metrics are included in Figure 12.1. For the item response analysis, data sets were constructed for each section using complete student records. Any student record that contained a missing score in any particular section was removed from that section’s analysis. For example, if a student had a missing score in Section A1, but had complete scores for A2, that student was removed from the A1 analysis but retained for the A2 analysis. The analysis for each section was conducted separately by grade level, but the results from all grade levels were evaluated together to understand how the section functions as part of the CFOL assessment.

In the item response model, the ability level is the latent trait that is being measured. Ability is described on a continuous scale centered at the average ability level, which is assigned a value of zero. The difficulty parameter identifies the ability level at which a student has an approximately 50% chance of responding at or above the corresponding level of response (see Figure 12.1). Overall, results from the item response analysis indicated that, as designed, items progressed in difficulty by the order in which they were placed in each section. Additionally, higher grade levels produced larger mean scores than lower grade levels, indicating that the measure adequately assessed a progression of skill level.

Figure 12.1
Item Response Analysis Metrics

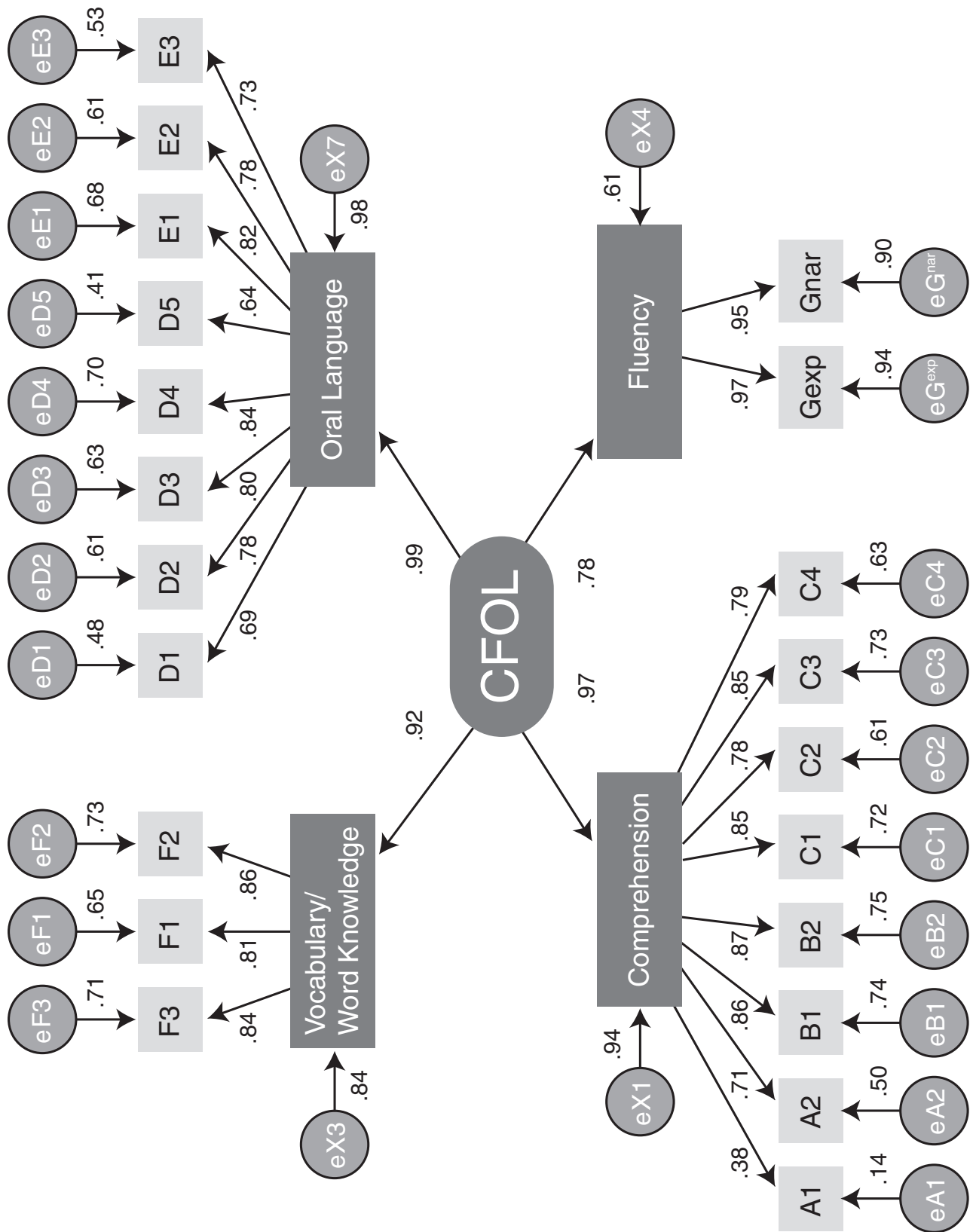
IRT Metric	Purpose for Evaluation of CFOL
Level of Response	Each section in the CFOL assessment is scored on a discrete scale that reflects a level of knowledge. Items within the same section had the same potential levels of response. For example, in Section A1, Storytelling, the levels of response ranged from 0–7 for each item. The percent of students that scored at each level was evaluated to assess (a) if there were any items that were significantly more difficult or easy than the other items within the same section, (b) if each item displayed a consistent score, and (c) if the items varied in difficulty.
Ability Level	In an item response model, the ability level is the latent trait that is being measured. Ability is described on a continuous scale centered at the average ability level, which is assigned a value of zero.
Difficulty	Difficulty parameters were evaluated to measure the ability level of each item. The difficulty parameter identifies the ability level at which a student has an approximately 50% chance of responding at or above the corresponding level of response. For example, in Section A1, Storytelling, the difficulty parameters for all items increase across the levels of response, and range from -2.3 to 1.7, indicating that student responses require higher levels of ability for higher levels of response, just as designed.
Discrimination	Discrimination estimates were evaluated to identify those items that do and do not work well across grade levels. The discrimination parameter explains how well an item can discriminate between those who scored at or above the difficulty parameter (highly proficient students) and those who scored below (less proficient students). For some sections, the discrimination estimates were very low on all grades. Those items were candidates for removal from the assessment.
Information and AUC	Information was evaluated to gauge the range of ability level for each item and section. The information function describes the precision with which the assessment measures each level of the underlying latent construct. Information is typically presented graphically where ability level is plotted on the x-axis and degree of precision is plotted on the y-axis. The area underneath the information curve (AUC) indicates the probability that student ability will be captured within a given range.
Cronbach's Alpha	Cronbach's alpha was calculated to evaluate the internal consistency and reliability of items within each individual CFOL section.

Confirmatory factor analysis. The analysis was conducted with SPSS® AMOS using maximum likelihood estimation. Student data from all grades were combined in a single model in order to capture the most variability at the task level. Tasks were separated into categories based on the skills they measure (comprehension, fluency, and oral language skills, including vocabulary), and the categories were connected by a single latent construct.

To assess model fit, the Akaike information criterion (AIC) from the confirmatory model was compared to the AIC from two comparison models: (a) a baseline model in which tasks were grouped together under their section categories (e.g., A1 and A2, B1 and B2, etc.) and (b) an alternate model with a different mix of tasks under the comprehension and fluency categories. The confirmatory model had the lowest AIC, indicating that it was the best model. The difference in AIC between the confirmatory model and the comparison models was 48.11 for comparison model 1 and 96.83 for comparison model 2, both of which are well above the threshold of 10 for significantly different model fit (Burnham & Anderson, 2002). The confirmatory factor analysis model is reported in Figure 12.2.

Figure 12.2

Factor Structure Map and Communality Estimates for Acadience Reading Diagnostic CFOL



Changes to CFOL based on the IRT and CFA. The results of the IRT for Listening Comprehension Retell (B1), Listening Comprehension Paragraph Question & Answer (B2), Paragraph Retell (C1), Sentences with Homophones (C2), Sentences with Homographs (C3), Passage Retell (C4), Matching Sentences to Pictures (D1), Use of Plurals (D2), Use of Past Tense (D3), Morpheme Compounding (E1), Multiple Meanings (F2), and both reading fluency passages in Section G indicated that the items measured increasing skills, represented a range of difficulty appropriate to the grade level, discriminated well between ability levels, and possessed good internal consistency reliability. These sections were not changed as a result of this analysis.

Patterns in the results from the IRT for Sentence Anagrams (D4), Sentence Repetition (D5), Sentence Completion (Decomposition, Derivation, & Inflection) (E2), Making Words (E3), Definitions (F1), and Figurative Language (Idioms) (F3) suggested that some items did not function as intended and could be removed. For Story Telling (A1), the distribution of scores across levels of the response, difficulty parameters, and the item information functions suggested that the scoring rubric could be condensed from a 7-point scale to a 5-point scale.

Three sections were removed based on the results from the IRT in combination with the results from the confirmatory factor analysis and correlational analysis. These three sections involved morpheme identification, syntax discrimination, and summarizing main ideas in short passages.

Reliability

Reliability refers to the relative stability with which a test measures the same skills across minor differences in conditions. Information about a test's reliability can be obtained in a variety of ways. In this study, we estimated reliability by examining (a) inter-rater reliability, using percent agreement and kappa; (b) internal consistency, using Chronbach's alpha; and (c) the communality estimates from the factor analysis. In this section, we report the results of the inter-rater reliability analysis first, followed by the Chronbach's alpha results and the communality results from the factor analysis.

Rater agreement. Salvia et al. (2007) suggest two approaches for evaluating the extent to which we can generalize to different scorers. The first approach is the correlational relationship between the scores of the two raters. The second approach is evaluating the proportional relationship of exact agreement between the raters. Salvia et al. suggest Cohen's kappa for the latter, which adjusts the proportion of agreement for the possibility of chance and is the lowest probable value for an estimate of rater reliability. In this report, we present both approaches side by side for comparison and correspondence. Hopkins (2006) provides guidelines for interpreting the strength of the correlation, and Landis and Koch (1977) provide guidelines for interpreting kappa. These guidelines are given in Table 12.19. Results for rater agreement are reported in Table 12.20.

The results indicate that the correlational relationship between raters is strong to almost perfect for all tasks, and ranges from .76 (E3. Making Words) to .99 (C2. Sentences with Homophones, C3. Sentences with Homographs, and E2. Sentence Completion (Decomposition, Derivation, & Inflection)). The proportional agreement relationship between raters (i.e., kappa) is moderate to almost perfect, and ranges from .41 (F1. Vocabulary Definitions) to .93 (E2. Sentence Completion (Decomposition, Derivation, & Inflection)). Kappa estimates and correlations were both strong for 15 of the 21 tasks indicating good correspondence with the correlational relationship between raters. Correlations were almost perfect (above .90) for 18 out of the 21 tasks, indicating that even when there was not exact agreement, the scores were very close. These results suggest very high reliability between raters.

Table 12.19
Rater-Agreement Reliability Estimate Descriptors

Kappa Range	Correlation Range	Descriptor
.81–1.0	.91–1.0	Almost Perfect
.61–.80	.71–.90	Strong
.41–.60	.51–.70	Moderate
.21–.40	.31–.50	Fair
0–.20	0–.30	Slight
< 0	< 0	Poor

Table 12.20
Rater Agreement for CFOL Sections

Section	Task	Rater Agreement			
		Grades	N	Kappa	Correlation
Story Coherence/Text Structure					
A1	Story Telling	K-1	39	.53	.95
A2	What Happens Next	K-1	21	.49	.93
Listening Comprehension					
B1	Retell	K-1	38	.46	.95
B2	Question and Answer	K-1	39	.57	.94
Reading Comprehension					
C1	Paragraph Retell	2	21	.67	.90
C2	Sentences with Homophones	2	23	.89	.99
C3	Sentences with Homographs	2-4	60	.90	.99
C4	Passage Retell	3-4	37	.44	.87
Syntactic Knowledge/Grammar					
D1	Matching Sentences to Pictures	K	21	.77	.95
D2	Use of Plurals	K-3	78	.76	.91
D3	Use of Past Tense	K-3	77	.85	.97
D4	Sentence Anagrams	K-3	77	.87	.94
D5	Sentence Repetition	K-4	98	.79	.95
Morphological Awareness					
E1	Morpheme Compounding	K-1	58	.89	.97
E2	Sentence Completion (Decomposition, Derivation, & Inflection)	K-4	97	.93	.99
E3	Making Words	1-4	77	.65	.76
Vocabulary/Word Knowledge					
F1	Definitions	K-4	92	.41	.84
F2	Multiple Meanings	K-4	91	.77	.97
F3	Figurative Language (Idioms)	1-4	75	.59	.92
Reading Fluency					
G	Expository Passage	1-4	51	.64	.97
G	Narrative Passage	1-4	50	.59	.96

Note: All correlations significant at $p < .001$.

Internal consistency. To assess the consistency of items, the results from two analysis were evaluated. First, an item-response analysis (IRT) was conducted on individual sections of the CFOL assessment. Second, the results from a confirmatory factor analysis was evaluated to verify the factor structure.

Chronbach's alpha. Cronbach's alpha was calculated for each section to assess internal consistency reliability of the items (i.e., the degree to which the items within each section measure similar skills). Grade-level and median alpha is reported in Table 12.21. Out of the 21 tasks, three

demonstrate acceptable internal consistency (above .60), 13 demonstrate good internal consistency (above .70), and the two fluency passages demonstrate excellent internal consistency (above .90). Three tasks demonstrate poor consistency (C2, D1, D3, E2), but suggest greater reliability in some grade levels over others (C2, D3, E2), and have median coefficients on the cusp of acceptable levels. Overall, these reliability estimates suggest that the items within the CFOL assessment possess good internal consistency reliability.

Table 12.21
Cronbach's Alpha Reliability Coefficients for CFOL Sections

Section	Task	Reliability (Cronbach's Alpha)					Median
		K	1	2	3	4	
A1	Story Telling	.82	.88	–	–	–	.85
A2	What Happens Next	.68	.63	–	–	–	.66
B1	Retell	.83	.80	–	–	–	.82
B2	Question & Answer	.80	.72	–	–	–	.76
C1	Paragraph Reading Retell	–	–	.86	–	–	.86
C2	Sentences with Homophones	–	–	.44	.55	–	.49
C3	Sentences with Homographs	–	–	.73	.80	.68	.73
C4	Passage Retell	–	–	–	.73	.84	.78
D1	Matching Sentences to Pictures	.49	.49	–	–	–	.49
D2	Use of Plurals	.50	.68	.66	.64	–	.65
D3	Use of Past Tense	.49	.65	.61	.49	.58	.58
D4	Sentence Anagrams	.84	.80	.71	.66	–	.76
D5	Sentence Repetition	.80	.79	.76	.70	.64	.76
E1	Morpheme Compounding	.64	.65	.67	–	–	.65
E2	Sentence Completion (Decomposition, Derivation, & Inflection)	.44	.40	.51	.61	.34	.44
E3	Making Words	–	.85	.84	.82	.81	.83
F1	Definitions	.81	.79	.79	.86	.88	.81
F2	Multiple Meanings	.88	.85	.86	.91	.86	.86
F3	Figurative Language (Idioms)	–	.76	.74	.78	.73	.75
G	Reading Fluency (Expository)	–	–	.89	.79	.91	.89
G	Reading Fluency (Narrative)	–	–	.90	.89	.90	.90

Note: $n = 475$.

Reliability derived from CFA communality estimates. Communality estimates indicate strong relationships between the observed scores and the latent construct, indicating that CFOL scores can be interpreted to appropriately measure reading comprehension. Reliability estimates were calculated from the communality estimates. Communality and reliability estimates for the skill categories and tasks are reported in Table 12.22.

Table 12.22
Skill Category and Section Reliability Estimates

Section	Task	Grade	Communality	Reliability
<i>Skill Category</i>				
	Comprehension	K–4	.97	.94
	Fluency	1–4	.78	.61
	Oral Language	K–4	.99	.98
	Vocabulary/Word Knowledge	K–4	.92	.85
<i>Subtest</i>				
Story Coherence/Text Structure				
A1	Story Telling	K–1	.38	.14
A2	What Happens Next (Prediction)	K–1	.71	.50
Listening Comprehension				
B1	Retell	K–1	.87	.76
B2	Question and Answer	K–1	.86	.74
Reading Comprehension				
C1	Paragraph Retell	2	.85	.72
C2	Sentences with Homophones	2–3	.78	.61
C3	Sentences with Homographs	2–4	.85	.72
C4	Passage Retell	3–4	.79	.62
Syntactic Knowledge/Grammar				
D1	Matching Sentences to Pictures	K–1	.69	.48
D2	Use of Plurals	K–3	.78	.61
D3	Use of Past Tense	K–4	.80	.64
D4	Sentence Anagrams	K–3	.84	.71
D5	Sentence Repetition	K–4	.64	.41
Morphological Awareness				
E1	Morpheme Compounding	K–2	.82	.67
E2	Sentence Completion (Decomposition, Derivation, & Inflection)	K–4	.73	.53
E3	Making Words	1–4	.74	.55
Vocabulary/Word Knowledge				
F1	Definitions	K–4	.86	.74
F2	Multiple Meanings	K–4	.81	.66
F3	Figurative Language (Idioms)	1–4	.84	.71
Reading Fluency				
G	Expository Passage	1–4	.97	.94
G	Narrative Passage	1–4	.95	.90

Note: Estimates were calculated from a confirmatory factor analysis. Sample size = 470. All grades were included in a single confirmatory factor analysis model. Goodness of fit statistics were not available due to nonrandom missing data (i.e., tasks are grade-level specific).

Consumer satisfaction

A link to the anonymous electronic consumer satisfaction questionnaire was sent to assessors at the conclusion of the study. Fifty-one assessors responded to the questionnaire. There were 14 items that were rated using a Likert-type scale with 1 = Strongly Disagree, 2 = Disagree, 3 = Somewhat Disagree, 4 = Somewhat Agree, 5 = Agree, and 6 = Strongly Agree. Responses were by and large quite positive. For example, for Item 2 (the administration and scoring rules were easy to follow), 60% of respondents indicated some level of agreement (i.e., either slightly agree, agree, or strongly agree) with this statement (see Figure 12.3). For Item 3 (the materials were organized appropriately for efficient administration of the measure(s)), 88% of respondents indicated some level of agreement (see Figure 12.4). Regarding Item 9 (the measures were a good way to assess students' reading strengths and weaknesses related to comprehension, reading fluency, oral language, and vocabulary) and Item 15 (overall, the measures would be beneficial for planning reading instruction for struggling readers), 75% and 74% of respondents indicated some level of agreement, respectively (see Figures 12.5 and 12.6).

Figure 12.3
Questionnaire Results for Item 2

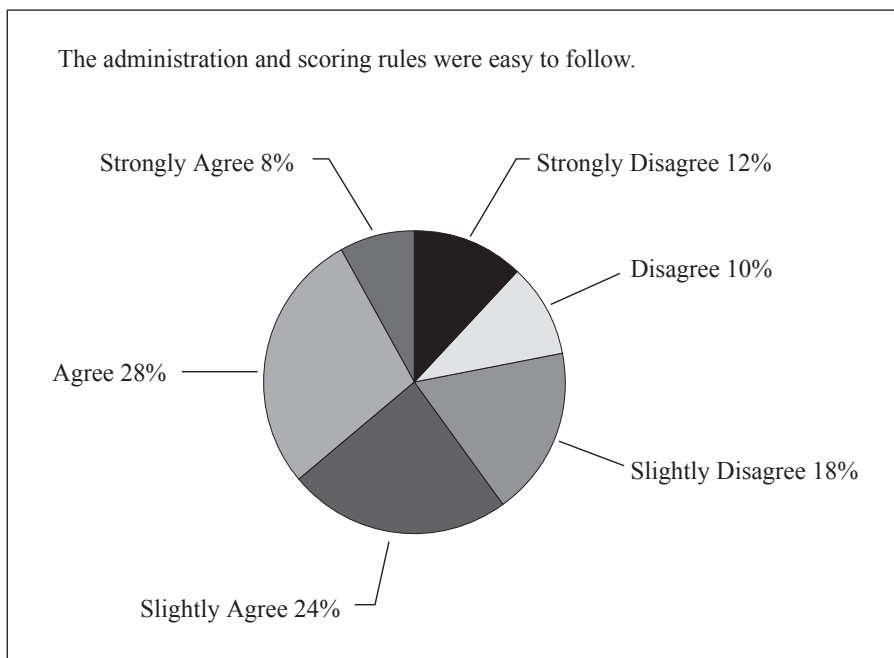


Figure 12.4
Questionnaire Results for Item 3

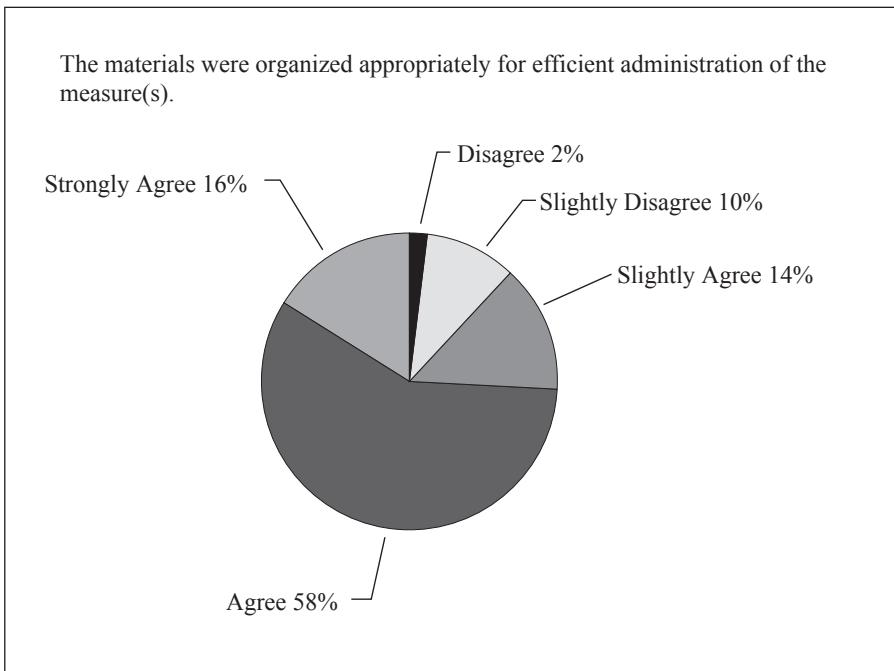


Figure 12.5
Questionnaire Results for Item 9

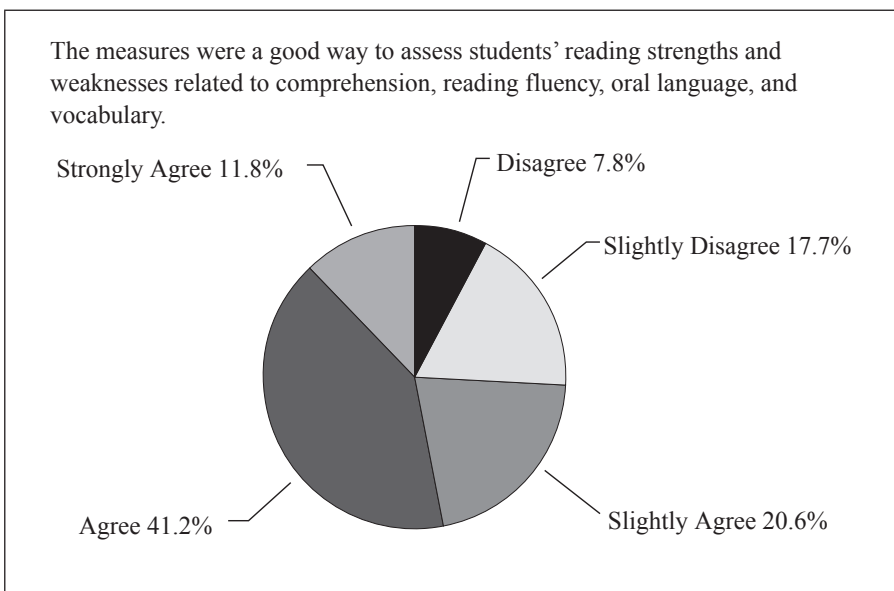
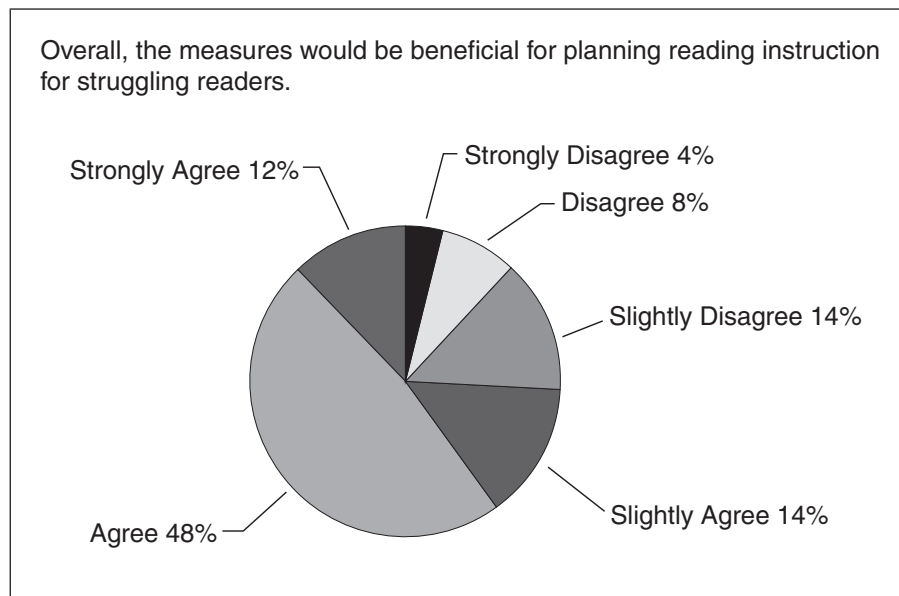


Figure 12.6
Questionnaire Results for Item 15



In addition to the responses rated on the Likert-type scale (from Strongly Agree to Strongly Disagree), assessors provided numerous comments. Taken together, the feedback provided was constructive but generally positive. Common concerns expressed were related to difficulty scoring certain sections (e.g., Making Words, Definitions, & Figurative Language) and the length of time required to carry out the testing in the context of this study (all items and no discontinue rules). All consumer feedback was carefully considered when revising the measures.

References

- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information–theoretic approach* (2nd ed.). Springer–Verlag.
- Good, R. H., Kaminski, R. A., Dewey, E. N., Wallin, J., Powell–Smith, K. A., & Latimer, R. J. (2013, revised 2019). *Acadience Reading K–6 technical manual*. Acadience Learning Inc.
<https://acadiencelearning.org/>
- Hopkins, W. G. (2006, August). *A scale of magnitudes for the effect statistics*.
<http://www.sportsci.org/resource/stats/effectmag.html>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 671–679.
- Powell-Smith, K. A., Dewey, E., Latimer, R., & Warnock, A. (2015). *Acadience Reading Diagnostic CFOL validation study technical report*. [Manuscript in preparation]. Acadience Learning Inc.
- Salvia J., Ysseldyke, J., & Bolt, S. (2007). *Assessment in special and inclusive education* (4th ed.). Houghton Mifflin Company.