2019–2020 LEAP Connect
Operational Technical Summary

English Language Arts, Mathematics, and Science

MEASUREMENT INCORPORATED

edCount LLC
because all students count

2019–2020 Technical Summary

The tests used in Louisiana are carefully constructed to fairly assess the progress of Louisiana students. This document provides an overview of the process and summarizes some of the key psychometric information for the LEAP Connect assessments.

Introduction

In December of 2016, the Louisiana State Board of Elementary and Secondary Education (BESE) approved new Louisiana Connectors (LCs) aligned to the 2016 Louisiana Student Standards (LSS) in ELA and mathematics. These connectors are designed for use in the instruction and assessment of students with significant cognitive disabilities. They are derived from the general education standards, but are reduced in depth, breadth, and complexity. The LCs in ELA and mathematics replaced what were formerly known as the Extended Standards. After the new LSS in science were approved in 2017, Louisiana began working with edCount, LLC, to develop LCs for science aligned to these new standards. The LCs for science were approved shortly after the adoption of the LSS for science.

In the 2017–2018 school year, Louisiana implemented the new LEAP Connect assessments in ELA and mathematics, which are fully aligned to the new LCs. The LEAP Connect assessments replaced the LAA1 assessment in ELA and mathematics, grades 3–8 and high school. The LEAP Connect assessments in ELA and mathematics for high school were first administered in the 2018–2019 school year.

The LAA1 science assessments were still used in 2017–2018 while the state worked with its vendor on the development of a new LEAP Connect science assessment aligned to the LCs in science. The science assessments were first administered in the 2019–2020 school year as census field tests. The first operational administration will take place in spring of 2021. The LEAP Connect science assessments will assess students in grades 4, 8, and high school. These are the same grades assessed by their predecessor, the LAA1 science assessments.

Louisiana's *Bulletin 111* §3901 states that all students, including those with disabilities, shall participate in Louisiana's testing program. To be eligible to participate in the LEAP Connect assessments, an IEP team must verify that the student has a disability which significantly impacts cognitive functioning and meets the criteria outlined in *Bulletin 1530* §505.

Federal law requires states to administer annual assessments to all students, including students with significant cognitive disabilities, to measure progress towards challenging academic content standards. The LEAP Connect assessments in ELA, mathematics, and science fulfill this requirement, in accordance with Sections 1111(b)(1)(E) and 8401 of the Elementary and Secondary Education Act of 1965.

Louisiana's *Bulletin 111* §3901 states that all students, including those with disabilities, shall participate in Louisiana's testing program. To be eligible to participate in the LEAP Connect assessments, an IEP team must verify that the student has a disability which significantly impacts cognitive functioning and meets the criteria outlined in *Bulletin 1530* §505. The LEAP Connect is designed for students with significant cognitive disabilities who cannot participate in the LEAP 2025 assessment, even with accommodations. Louisiana's *Bulletin 111* §3901 states that all students, including those with disabilities, shall participate in Louisiana's testing program. To be eligible to participate in the LEAP Connect assessments, an IEP team must verify that the student has a disability which significantly impacts cognitive functioning and meets the criteria outlined in *Bulletin 1530* §505. The LEAP Connect is

designed for students with significant cognitive disabilities who cannot participate in the LEAP 2025 assessment, even with accommodations.

Louisiana believes that all students, including those with the most significant cognitive disabilities, deserve an education that prepares them to be independent and successful in life after high school.

The Student Population

The LEAP Connect assessment system is designed for students with significant cognitive disabilities for whom participation in the general assessments would not be appropriate, even with accommodations. Understanding the characteristics of this population is a vital aspect of maintaining an effective system of instruction and assessment and ensuring the system is serving the appropriate population. The Louisiana students who participate in the LEAP Connect must meet the following criteria:

1. The student has a disability that significantly impacts cognitive function and/or adaptive behavior.
2. The student requires extensive modified instruction aligned with the Louisiana Connectors to acquire, maintain, and generalize skills.
3. The decision to include the student in the alternate assessments is not solely based on certain factors (placement, behavior, English Learner status, etc.).

Section 1111(b)(2)(D)(i)(I) of the Elementary and Secondary Education Act of 1965 (ESEA), as amended by the Every Student Succeeds Act (ESSA), states that no more than 1% of a state's total student population may participate in the alternate assessments. Louisiana exceeded this cap in the past few years in ELA and mathematics. The state did not exceed the 1% cap in science. The LDOE was granted a waiver for the 2017–2018 and 2018–2019 school years. However, the waiver for the 2019–2020 school year was denied.

As part of the effort to meet the 1% cap requirement, the LDOE has required each local education agency (LEA) that exceeds the 1% cap to:

- Provide written justification describing the specific reason(s) the percentage of students taking the alternate assessments exceeds 1%;
- Provide written assurance that the LEA followed the state's guidelines for participation in the alternate assessments; and
- Provide written assurance that the LEA would address any disproportionality in the percentage of students in any subgroup taking an alternate assessment.

In addition, the LDOE revised the alternate assessment eligibility criteria and deployed accountability and transparency enhancements to the statewide IEP system. The LDOE has provided additional resources and support to LEAs and educators to assist with implementing these changes, including but not limited to:

- Training and support to LEAs to clarify the revised eligibility criteria;
- A new webpage dedicated to students with significant cognitive disabilities;
- A resource library for students with significant cognitive disabilities;
- Individualized support for LEAs whose student-level files indicated that IEP team decisions were not consistent with state participation criteria.

The participation rates for the 2017–2018, 2018–2019, and 2019–2020 school years are provided in Exhibit 1. Both the percentage of students with significant cognitive disabilities participating in the LEAP Connect out of all students eligible to participate in this assessment and the percentage of students with significant cognitive disabilities assessed via the LEAP Connect out of the entire Louisiana student population are presented.

Exhibit 1. Alternate Assessment Participation Rates

	2017–2018		2018–2019		2019–2020	
Content Area	**% of Eligible SWSCD**	**% of All LA Students**	**% of Eligible SWSCD**	**% of All LA Students**	**% of Eligible SWSCD**	**% of All LA Students**
ELA	99.0	1.3	98.8	1.6	98.4	1.5
Math	98.8	1.3	98.7	1.6	98.3	1.5
Science[1]	98.9	0.7	97.8	0.7	100.0	0.7

Test Content Development

The LEAP Connect assessments measure student proficiency and achievement in ELA and mathematics in grades 3–8 and high school, and in science in grades 4, 8, and high school. The LEAP Connect system assesses student proficiency in terms of the Louisiana Connectors (LCs), which are fully aligned to the Louisiana Student Standards (LSS) for ELA, mathematics, and science. Each assessment provides age- and grade-appropriate content for all grades and courses while maintaining high expectations for all students, capturing the "big ideas" found in the LSS.

The LCs are utilized for assessment purposes in that they were designed to reflect the necessary knowledge and skills that students with the most significant cognitive disabilities need to reach critical learning targets or big ideas within the standards from grade band to grade band, leading to knowledge of ELA, mathematics, and science for college, career, and community readiness by the end of high school.

The LCs are designed to provide fully aligned pathways for students with significant disabilities to work toward the LSS. The LCs identify the:

- Most salient grade-level, core academic content found in the LSS;
- Necessary knowledge and skills needed to reach grade-level expectations of the LSS;
- Core content, knowledge, and skills needed at each grade to promote success at the next; and
- Priorities in each content area to guide the instruction for students in this population.

Principled Design and Universal Design

The LEAP Connect assessment system was designed according to the principles of principled design and Universal Design (UD). According to AERA et al. (2014, pp. 6–7), tests should be designed to minimize construct-irrelevant barriers for all test takers in the target population. Thus, an understanding about student characteristics and the application of UD principles inform the design of each item and any

[1] Reflects LAA1 Science participation in 2017–2018 and 2018–2019 and the LEAP Connect census field test participation in 2019–2020.

necessary additional adaptations and accommodations that do not interfere with the measured construct. The principled design approach focuses the development of items for all students on construct-relevant content (i.e., the knowledge, skills, and abilities intended to be measured), minimizing the impact of construct-irrelevant skills (e.g., print size, lack of assistive technology device, inability to engage with the items), and considering appropriate accessibility options.

The definition and implementation of accessibility features for all aspects of the assessment development process to provide universal access (beyond what is currently achieved through accommodations and Universal Design) is necessary to support improved performance for English Learners (ELs), students with disabilities, students with 504 plans, and students with disabilities who are ELs (Almond et al., 2010).

To this end, the LEAP Connect assessment developers incorporated the guidelines of UD as described by the National Center on Universal Design for Learning (http://www.udlcenter.org/). Developers addressed the vast majority of student access needs (e.g., cognitive, processing, sensory, physical, language) up front in the design of the assessment items. This was done by embedding specific accessibility features (e.g., magnification, audio representation of graphic elements, linguistic simplification) into the structure and delivery of the assessment items and formats.

The LEAP Connect assessments administered February 3 to March 13, 2020, are fixed-form, computer-based tests administered online through the DRC INSIGHT platform. They are administered in a one-to-one setting and include both selected-response and constructed-response items. The assessments include several features that promote accessibility, including:

- The entire test can be read aloud to students.
- Students may respond to items based on their preferred mode of communication (e.g., eye gaze, assistive technology, point to a picture, etc.).
- Items include pictures and graphics to support what is read to students. Nearly all of the mathematics items contain visual stimuli to assist students in determining an answer.
- Items indicate when students may use calculators. Any student with an IEP accommodation for calculator use may use their specified calculator for every item. While an online calculator is provided, students may use the handheld calculator they typically use during instruction on the mathematics test.

Test Specifications

The LEAP Connect assessment items are written based on common item and test specifications, which establish performance levels with achievement level descriptors for ELA, mathematics, and science. The test specifications for the LEAP Connect assessments for ELA, mathematics, and science provide general guidelines for the development of all test items used in the assessments for each content area and grade level.

The assessment blueprints, as part of the overall test specifications, provide valid information about students' knowledge and skills in ELA, mathematics, and science in relation to the LCs. The blueprints also define what is centrally important, represent a balance of emphasis, and are vertically sequenced. For each content area, the LEAP Connect assessment blueprints include the content category, weight (as a percentage), LC, item type (selected-response or constructed-response), and number of score points for each assessed grade.

LEAP Connect items are written at four levels of complexity. To access the age- and grade-appropriate general curriculum content and to build skills and knowledge in ELA, mathematics, and science, students with significant cognitive disabilities often need adaptations, scaffolds, and supports. For students to accurately demonstrate what they know and can do, these age- and grade-appropriate adaptations, scaffolds, and supports also need to be present within the assessment process. The assessment items incorporate important aspects of item design related to both varying levels of cognitive complexity and the degree and type of scaffolds and supports.

Reliability

The reliability of raw scores by test form was evaluated using Cronbach's (1951) coefficient alpha, which is a lower-bound estimate of test reliability. The reliability coefficient is a ratio of the variance of true test scores to the variance of the total observed scores, with the values ranging from 0 to 1. The closer the value of the reliability coefficient is to 1, the more consistent the scores, where 1 refers to a perfectly consistent test. In general, reliability coefficients that are equal to or greater than 0.8 are considered acceptable for tests of moderate length.

The reliability of reported test scores can be characterized by the standard errors associated with the scores. The standard error of measurement (SEM) may be used to determine the range within which a student's true score is likely to fall. An observed score should be regarded not as a student's true score but as an estimate of a student's true score. It is expected that the score a student obtains from a single test administration would fall within one SEM of the student's true score 68% of the time and within approximately two SEMs of the true score 95% of the time.

Total test reliability measures, such as Cronbach's coefficient alpha and SEM, consider the consistency (i.e., reliability) of performance over all test questions in a given form, the results of which imply how well the questions measure the content domain and could continue to do so over repeated administrations. The number of items in the test influences these statistics; a longer test can be expected to be more reliable than a shorter test.

The reliability coefficients and SEM for the LEAP Connect assessments are reported in Exhibit 2. The reliability statistics ranged from 0.82 to 0.87 for the ELA forms. For mathematics, the reliabilities ranged from 0.77 to 0.87. For science field tests, the reliability values were 0.85 or above for all the forms and grades. These results indicate acceptable reliability coefficients for the LEAP Connect assessments.

Exhibit 2. Reliability and SEM

Content Area	Grade	N Items	Max Score Points	SEM	Cronbach's Alpha	N-Count
ELA	3 NV	29	30	2.70	0.87	>=220
	3 V	29	30	1.95	0.86	>=300
	4 NV	30	31	2.79	0.86	>=190
	4 V	30	31	2.13	0.85	>=390
	5	29	30	2.81	0.82	>=660
	6	29	30	2.50	0.87	>=900
	7	28	29	2.46	0.86	>=950
	8	30	31	2.44	0.85	>=1040
	11	27	28	2.35	0.87	>=980
Math	3	35	35	3.11	0.87	>=520
	4	33	33	2.87	0.85	>=580
	5	35	35	3.30	0.77	>=660
	6	35	35	2.99	0.85	>=900
	7	34	34	2.81	0.86	>=950
	8	35	35	2.84	0.87	>=1040
	11	35	35	2.92	0.87	>=990
Science	4 F1	30	30	2.70	0.87	>=290
	4 F2	30	30	2.63	0.88	>=280
	8 F1	30	30	2.55	0.85	>=510
	8 F2	30	30	2.62	0.85	>=510
	HS F1	30	30	2.53	0.89	>=470
	HS F2	30	30	2.54	0.88	>=470

Construct Validity

In addition to content validity addressed in the Test Content Development and Reliability sections, additional evidence of validity, especially construct-related validity, is demonstrated through studies of convergent and divergent validity.

Convergent validity is a subtype of construct validity that can be estimated by the extent to which measures of constructs that theoretically should be related to each other are, in fact, observed as related to each other. Analyses of the internal structure of a test can indicate the extent to which the relationships among test items conform to the construct the test purports to measure.

Divergent validity is a subtype of construct validity that can be assessed by the extent to which measures of constructs that theoretically should not be related to each other are, in fact, observed as not related to each other. Typically, correlation coefficients among measures of unrelated or distantly related constructs are examined in support of divergent validity.

Minimization of construct-irrelevant variance and construct underrepresentation is addressed in the following steps of the test development process: (1) specification, (2) item writing, (3) review, (4) field testing, (5) test construction, and (6) item calibration.

Construct-irrelevant variance refers to error variance that is caused by factors unrelated to the constructs measured by the test. For example, when tests are not administered under standardized conditions (e.g., one administration may be timed, but another administration is untimed), differences in student performance related to different administration conditions may result. Careful specification of the content and the review of the items representing that content are first steps in minimizing construct-irrelevant variance. Then, empirical evidence, especially item-level data, is used to infer construct irrelevance.

Construct underrepresentation occurs when the content of the assessment does not reflect the full range of content that the assessment is expected to cover. Specification and review, a process through which test blueprints are developed and reviewed, are primary steps in the development process designed to ensure that content is appropriately represented.

To present evidence of construct-related validity, the 2019-2020 LEAP Connect Technical Report describes in detail the following validity studies:

- Decision Accuracy and Consistency;
- Dimensionality and Local Independence;
- Divergent (Discriminant) Validity; and
- Item-Total Correlations.

In addition, the technical report addresses validity evidence based on the five factors specified in The *Standards for Educational and Psychological Testing* (*Standards*; AERA, APA, & NCME, 2014). These include:

- Test content;
- Response processes;
- Internal structure;
- Relationships to other variables; and
- Consequences of testing.

Uses of Test Scores

To understand whether a test score is being used properly, one must understand the purpose of the test. The intended uses of the LEAP Connect test scores include the following:

- To gauge student progress in relation to grade-level academic standards;
- To help educators improve their teaching practices year to year to raise student achievement; and
- To inform school accountability decisions.

LEAP Connect scores are not meant to be diagnostic in nature and are not used to alter instruction in real time. Rather, they provide an end-of-year snapshot that stakeholders at the state, district, school, and classroom levels can use to make informed decisions for the following school year.

Test Level Scores

At the test level, an overall scale score that is based on student performance on the entire test is reported. In addition, an associated level of achievement is reported. These scores and achievement levels indicate, in varying ways, a student's achievement in ELA, mathematics, or science. Test-level scores are reported at four reporting levels: the state, the school system, the school, and the student.

Two types of test-level scores are reported to indicate a student's achievement on the LEAP Connect: (1) the scale score and (2) its associated level of achievement.

Scale Scores

Scale scores are derived from raw scores (i.e., the number of items answered correctly). Raw scores depend on the items in a particular form of a test and can only be interpreted in terms of that particular set of test questions. This does not allow year-to-year or form-to-form comparison. Scale scores are more meaningful than raw scores because they maintain their meaning year-to-year, thus allowing comparisons of different test forms across the entire range of the ability scale. For LEAP Connect ELA and mathematics assessments, scale scores range from 1200 to 1290 for all grades.

Scale scores are used to represent student performance on LEAP Connect tests. A higher scale score represents more knowledge, skill, and ability than a lower scale score. Scale scores for the same test can be compared regardless of when students were tested or which form was taken. Scale scores are also averaged together to represent the overall performance of a school, a school system, and the state.

Scale scores are comparable for results within the same grade and the same content area across years.

Achievement Levels

Achievement levels describe how students perform based on Louisiana's expectations and how prepared they are for the next level of study. Achievement Level Descriptors for ELA and mathematics at grades 3–8 and 11 were developed through an iterative process involving multiple stakeholder groups. The achievement levels for each grade summarize the knowledge, skills, and abilities that the average student demonstrates within each scale-score range. Each achievement level is cumulative, meaning each level includes the knowledge, skills, and abilities of the preceding levels.

The number and percent in achievement levels are reported at the school, school system, and state levels. Since this information is based on scale scores, it is comparable across groups for the same test regardless of when the test was taken or which form was taken. Unlike scale scores, it may be used to monitor group performance over time, across grade levels.

Uses of Test Scores

The LEAP Connect scale scores and achievement levels provide summary evidence of student performance relative to the LSSs and LCs. Classroom teachers may use these scores as evidence of student achievement in the assessed content areas. At the aggregate level, school system and school administrators may use this information for activities such as curriculum planning. The results presented in the technical report provide evidence that the scale scores and achievement levels are valid and reliable indicators of what students know, understand, and are able to do relative to the LSSs and LCs in each content area.

Equating of Test Forms

For the 2020 administration year, the ELA and mathematics forms, originating from the National Center and State Collaborative (NCSC) of which Louisiana was a member, were re-administered except for mathematics grade 7 and high school. The NCSC item parameters were estimated using the two-parameter logistic (2PL) model. For details regarding the methodology applied, see the National Center and State Collaborative 2015 Operational Assessment Technical Manual (NCSC, 2016).

For the mathematics grade 7 and high school tests, new items without the existing NCSC item parameter estimates were added to the operational tests. Those new item parameters were estimated with the 2PL IRT model using the 2020 data and linked to the NCSC item pool.

Following NCSC calibration procedures, the two-parameter logistic model (2PL) was utilized to calibrate dichotomous items using the PARSCALE software with the calibration statement: CAL GRADED, LOGISTIC, CYCLE=(100,1,1,1,1), TPRIOR, SPRIOR; as mentioned in the NCSC technical manual (NCSC, 123).

The 2PL model is defined as follows:

$$P_i\left(X_i = 1 \middle| a_i, b_i, \theta_j\right) = \frac{\exp(Da_i\left(\theta_j - b_i\right))}{1 + \exp(Da_i\left(\theta_j - b_i\right))}$$

where X_i indexes the scored response on an item, i indexes the items, j indexes students, a_i represents item discrimination, b_i represents item difficulty, ϑ_j represent the student ability measure, and D is a normalizing constant equal to 1.701.

Matrix calibration was performed to estimate the item parameters. With this approach, source data sets (from 2018 to 2020) are processed such that responses to a particular item, regardless of forms and years where the item was administered, will occupy a single column in the calibration data set. Each row belongs to a student taking a specific form in one administration year. Columns related to items that were not taken by the student were regarded as missing observations. The calibration data set had all items administered to all students in years 2018, 2019, and 2020 for each content area and grade combination.

Item parameter estimation was performed without fixing any known (NCSC) parameters, which is commonly known as a free-run calibration. This procedure yields parameter estimates for all items administered between 2018 to 2020 administration years.

Equating

To build the score tables for mathematics grade 7 and high school, the 2020 forms were equated to match the original NCSC parameter scale. The Stocking and Lord equating procedure was used for this purpose.

For each content area and grade, the Stocking and Lord equating procedure starts by identifying the set of equating items. The initial set of equating items was given by the set of items with existing NCSC parameter estimates (anchor items). The numbers of anchor and equating items for each grade are given in Exhibit 3. PARSCALE item fit statistics were then used to exclude items that did not fit the data very well. Items were excluded from the equating set when the Chi-square probability is less than 0.05.

Exhibit 3. Number of Equating Items

Content Area	Grade	# Anchor Items	# Equating Items
Mathematics	07	28	14
	HS	21	19

The Stocking and Lord equating procedure derives the equating constants A and B, such that the summation of the squared difference between the total characteristic curves of equating items is minimized. That is,

$$SL_{diff} = \sum_{j} \left(\sum_{i \in V} P_i\left(X_i = 1 \middle| \widetilde{a_i}, \widetilde{b_i}, \theta_j\right) - \sum_{i \in V} P_i\left(X_i = 1 \middle| \frac{a_i}{A}, Ab_i + B, \theta_j\right) \right)^2$$

where V is the set of equating items, a_i and b_i are the discrimination and difficulty parameters from the free-run calibration, $\widetilde{a_i}$ and $\widetilde{b_i}$ are the NCSC discrimination and difficulty parameters, $\theta_j \in \{-4, -3.9, -3.8, \ldots, 3.8, 3.9, 4\}$. Theta values are chosen within the NCSC theta range of -4 to 4 (NCSC technical manual pg. 127).

The Stocking and Lord procedure performed very well in matching the test characteristic curve (TCC) of the equating items.

Results Summary

Results for ELA and Mathematics are summarized in Exhibit 4. Because the Science administration was a census field test, results cannot be summarized.

Exhibit 4. Results Summary

			Percent by Achievement Level					Scale Score	
Content Area	Grade	N	Level 1	Level 2	Level 3	Level 4	Levels 3 & 4	Mean	Std Dev
ELA	3	>=520	30.23	16.54	35.36	17.87	53.23	1239.14	17.39
	4	>=580	34.07	16.70	36.80	12.44	49.24	1240.54	17.86
	5	>=660	17.85	24.96	39.79	17.40	57.19	1241.83	17.60
	6	>=900	24.47	25.80	26.80	22.92	49.72	1241.95	18.16
	7	>=940	24.89	11.81	29.11	34.18	63.29	1246.69	18.44
	8	>=1030	16.14	28.02	22.03	33.82	55.85	1243.83	17.51
	HS	>=970	22.06	17.88	35.55	24.51	60.06	1246.23	18.49
Mathematics	3	>=520	30.40	14.91	31.93	22.75	54.68	1241.60	19.79
	4	>=580	25.39	16.81	30.36	27.44	57.80	1242.61	19.14
	5	>=660	17.85	33.13	29.50	19.52	49.02	1241.22	17.53
	6	>=900	22.67	20.11	22.44	34.78	57.22	1245.05	18.38
	7	>=940	10.97	20.04	33.12	35.86	68.98	1249.20	19.63
	8	>=1030	16.81	21.45	24.35	37.39	61.74	1246.54	18.43
	HS	>=990	21.37	22.68	22.68	33.27	55.95	1243.92	18.22

References

Almond, P., Winter, P., Cameto, R., Russell, M., Sato, E., Clarke-Midura, J., Torres, C., Haertel, G., Dolan, R., Beddow, P., & Lazarus, S. (2020). Technology-Enabled and universally designed assessment: Considering access in measuring the achievement of students with disabilities—A foundation for research. *Journal of Technology, Learning, and Assessment*, *10*(5). Retrieved July 9, 2021, from http://www.jtla.org.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards of Educational & Psychological Testing*.

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

National Center and State Collaborative (2016). *National Center and State Collaborative 2015 Operational Assessment Technical Manual.*