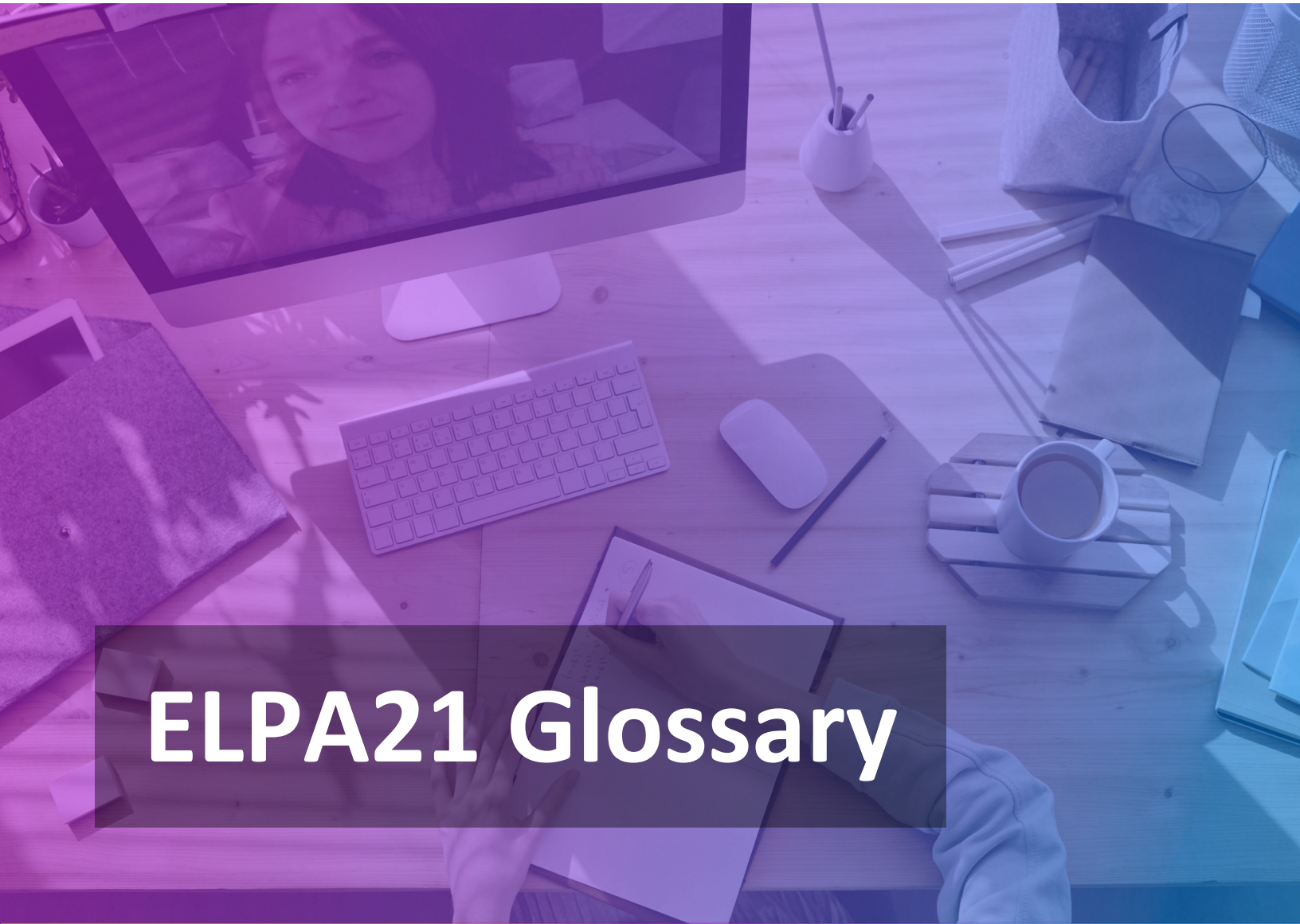![ELPA21 logo — English Language Proficiency Assessment for the 21st Century]

ELPA21 Glossary

English Language Proficiency Assessment for the 21st Century

March 2024

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

English Language Proficiency Assessment for the 21st Center (ELPA21)

ELPA21 is a comprehensive assessment and instructional system supporting the growth of educators and English learners. Through a strong partnership with educators and state learners, ELPA21 supports English learners in mastering language skills for academic success and provides educators with engaging professional learning and innovative instructional resources. ELPA21 exceeds standards for accessibility and accommodations for inclusive testing and instruction. The following document includes assessment and instruction terms related to ELP assessment and instruction.

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

ELPA21 Glossary of Terms

Accessibility: As used in educational assessment, accessibility refers to the degree to which the items or tasks on a test enable as many test takers as possible to demonstrate their understanding of the target construct without being impeded by characteristics of the item that are irrelevant to the construct being measured.[1]

Accessibility features: Embedded and non-embedded supports available to all students during a computer-based test. For example, allowing students to adjust the background color or contrast of the screen. Oftentimes, educators have to activate specific accessibility features prior to a test, based on a student's personal needs profile. Accessibility features are designed and intended to be used to ensure students receive equitable access to tests without the distraction of features they don't need.[2]

Accessible Portable Item Protocol (APIP): The Accessible Portable Item Protocol (APIP) Standard provides assessment programs and item developers with a data model for standardizing the interchange file format for digital test items. The APIP standard accomplishes two important goals. First, the standard allows digital tests and items to be ported across APIP compliant test item banks. Second, it provides a test delivery interface with all the information and resources required to make a test and an item accessible for students with a variety of disabilities and special needs.[3]

Accommodations: A change in administration of an assessment that increases equitable access for students with Individualized Education Programs (IEPs) or 504 plans. IEP teams and educators for 504 plans make the decisions about these accommodations. For English learners with disabilities, these teams should include an expert in second language acquisition. Accommodations include, but are not limited to, a change in assessment setting, scheduling, timing, presentation format, response mode, or any combinations of these changes, that does not change the construct intended to be measured by the assessment or the meaning of the resulting scores. For the complete list of accommodations available with the ELPA21 assessment system, refer to the ELPA21 Accessibility and Accommodations Manual.

Accountability: Accountability is the assignment of responsibility for conducting activities in a certain way or producing specific results. A primary motivation for increased accountability is to improve the system or aspects of it.[4]

Accountability system: A system that uses assessment results and other data that reflect the goals and expectations for students, teachers, schools, districts, and states to demonstrate the established components or requirements of accountability.

---

[1] AERA/APA/NCME Standards
[2] https://www.aps.edu/assessment/parcc-documents-folder/parcc-glossary-2-8-15
[3] http://www.imsglobal.org/APIP/index.html
[4] https://nceo.info/standards_and_accountability/accountability

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Achievement Level Descriptor (ALD): ALDs, also known as Performance Level Descriptors (PLDs), describe performance on the ELPA21 assessment as determined by the process of standard setting. ALDs are distinct from Achievement Level Indicators (ALIs), which describe expectations for English language proficiency as described by the ELP Standards.

Achievement Level Indicator (ALI): ALIs describe expectations for English language proficiency as described by the ELP Standards. ALIs are distinct from Achievement Level Descriptors (ALDs) or Performance Level Descriptors (PLDs), which describe performance on the ELPA21 assessment as determined by the process of standard setting.

Achievement test: Achievement tests are designed to measure the knowledge and skills that students learned in school or to determine the academic progress they have made over a period of time. The tests may also be used to evaluate the effectiveness of a school and its teachers, or identify the appropriate academic placement for a student—e.g., what courses or programs may be deemed most suitable, or what forms of academic support they may need. Achievement tests are "backward-looking" in that they measure how well students have learned what they were expected to learn.[5]

Age appropriate: The characteristics of the skills taught, the activities, and materials selected, and the language level employed that reflect the chronological age of the student.[6]

Aggregate score report: ELPA21 will provide states with aggregate score reports to school administrators and educators summarizing their students' performance grade-level and other groups as determined by state policy. Aggregate reports may contain the following subgroups:
1) EL status
2) Ethnicity
3) Gender
4) IEP status
5) Accommodations provided and/or used for testing
6) Title I status
7) Interrupted formal education status
8) Years of English language instruction
9) Native language

These reports provide aggregated scores at the school, district, and state level. States may add additional reports as necessary to describe regions or other institutional groupings specific to their state.

Aggregated data: Data combined from several measurements and expressed in a summary form, for purposes such as statistical analysis. A common aggregation purpose is to get more information about particular groups based on specific variables.

---

[5] http://edglossary.org/standardized-test/
[6] Ibid

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Alignment: The match or association between two or more elements or among content standards, performance standards, curriculum, instruction, and assessments in terms of knowledge and skill expectations.[7]

Alternate Assessment (AA): An assessment based on alternate standards developed for students with significant cognitive disabilities and for whom the general assessment, even with accessibility features, is not appropriate.[8]

Alternative and Augmentative Communication (AAC): A term used to describe the different methods that can be used to help people with disabilities communicate with others. These methods can be used as an alternative to speech or to supplement it and can include individual methods of sign and gestures, standardized signing, symbol systems, and complex electronic devices.[9]

Analysis of Variance (ANOVA): A collection of statistical techniques for comparing the results of more than two groups of students taking a test or to evaluate the amount of growth over time achieved by students.

Annual progress: The Every Student Succeeds Act requires that each state determine progress measures on an annual basis.

Answer key/Scoring key/Key: Specifications of the correct answer or range of correct/acceptable responses (where multiple responses are permitted) for an objective response test item.

Artificial Intelligence (AI) scored responses: Responses scored by software using one or more techniques to simulate the process and criteria used by human scorers. AI scoring is designed to generate scores consistent with those that would be produced by human scorers.[10]

Assessment: The wide variety of methods that educators use to evaluate, measure, and document the academic readiness, learning progress, and skill acquisition of students from preschool through college and adulthood. (See edglossary.org/assessment for more).[11]

Assessment literacy: Knowledge of the basic principles of sound assessment practice, including terminology, development, administration, analysis, and standards of quality.[12]

Assessment target: Level 4 on each domain represents the English language knowledge, skills and abilities that are required to interact with and engage in grade-level content instruction at the same level as non-ELs and is referred to as the "assessment target" for each domain. Once the assessment target is met on all non-exempt domains (e.g., a student scores "4444" (a four on each domain), ELPA21 recommends the student be eligible for reclassification.

---

[7] Ibid
[8] CCSSO's Operational Best Practices 2013
[9] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja
[10] CCSSO's Operational Best Practices 2013
[11] http://edglossary.org/assessment/
[12] Assessing Students with Disabilities: A Glossary of Assessment Terms in Everyday Language

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Assistive technology: Any software or device, for use by a student with a disability, that provides the student with enhancements to, or changes methods of interacting with, the technology needed to facilitate participation in an assessment.[13]

Attempted domain test: A domain test is "attempted" once the student has started the test (had the opportunity to view at least one item). A domain test is "not attempted" if the student never started the domain test (i.e., the student never had the opportunity to view any items).

Augment: To change test content, by adding, deleting, or customizing test items within an existing instrument to align that instrument with content standards.[14]

Authentic assessment: Authentic assessment is usually characterized by tasks and items that require students to construct a response or demonstrate skills and concepts in a real work context. Students usually develop responses, in writing or in performances, e.g., investigations, presentations, and portfolios, rather than by selecting predetermined options such as multiple-choice tests.[15]

Baseline data: The initial performance results, against which future results will be compared.[16]

Benchmarks: See *Learning Objectives.*

Bias: In a statistical context, a systematic error in individual or group test scores based on gender, ethnicity, disability, or other cultural dimensions. In discussing test fairness, bias may refer to construct under-representation (e.g., a lack of test items that are culturally sensitive) or the presence of construct-irrelevant components of a test score such as heavy reading emphasis on a math computation test.[17]

Blueprint: Blueprints are a series of documents that together describe the content and structure of a test. This often includes the subareas covered by the test and the number and type of questions included.

Body of Evidence (BOE): An array of information or data produced by a student or collected by someone who is knowledgeable about the student to establish that the student can perform one or more particular skills related to a content standard. The information or data in the BOE is judged against predetermined achievement standards.[18]

Calculating growth indicators for tests with exempted domains: If a student is exempted from taking a domain test, the Overall and Comprehension Scores for that student will be calculated based on a scoring model for the domains that were tested. Similarly, the profile of domain performance levels

---

[13] CCSSO's Operational Best Practices 2013
[14] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja
[15] Ibid
[16] Ibid
[17] http://www.edglossary.org/test-bias/
[18] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

will be evaluated for overall proficiency based on the domains tested. In other words, the student will not be "penalized" by treating the exempted domain as if the student had gotten all the items wrong.

Calculating growth indicators for tests with missing domains: If the student was supposed to take the reading or listening test (was not exempted) but did not, the student's Comprehension Score will be set to "N." A Comprehension Score based on a single domain is the same as the score for that domain. Since no additional information is added, no additional score will be reported. The Overall Score is based on all the items across the domains. The items from the missing domain will be treated as missing individual item responses.

Calibration: Is setting or establishing through Item Response Theory (IRT) methods, the parameters (e.g., difficulty, discrimination) of a series of items using student responses.

Claim: A statement used in ELPA21 item development that describes student performance within each domain based expected student performance within each domain. Claims come from the ELP standards and are paired with evidence statements that describe how each claim made will be supported and demonstrated by student responses.[19]

Cognitive laboratory: A session in which students are observed taking test items and are then asked to explain how they arrived at their responses. The students' explanations supplement the statistical evidence gathered about the items, testing tools, and testing systems. In general, a cognitive laboratory occurs prior to operational use of the items.[20]

College and Career Readiness (CCR): College and career readiness refers to the content knowledge, skills, and habits that students must possess to be successful in postsecondary education or training that leads to a sustaining career. A student who is ready for college and career can qualify for and succeed in entry-level, credit-bearing college courses without the need for remedial or developmental coursework.[21]

Common Core State Standards (CCSS): The Common Core is a set of high-quality, internationally benchmarked academic standards in mathematics and English language arts/literacy (ELA). These learning goals outline what a student should know and be able to do at the end of each grade. The standards were created to ensure that all students graduate from high school with the skills and knowledge necessary to succeed in college, career, and life. Forty-three states, the District of Columbia, four territories, and the Department of Defense Education Activity (DoDEA) have voluntarily adopted and are moving forward with the Common Core.[22]

Comparability study: A study comparing two modes of administration of a test, such as a comparison of a test administered via computer versus a test administered via paper and pencil.

---

[19] Dr. Phoebe Winter, Consultant
[20] CCSSO's Operational Best Practices 2013
[21] https://eric.ed.gov/?id=ED537876
[22] https://www.inflexion.org/category/inflexion/page/8/

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Compensatory model: A compensatory scoring model allows for high performance in one domain to compensate for low performance in another domain when determining overall proficiency. Using an average of domain scores to determine overall proficiency would be a compensatory model.

Complexity: The level of cognitive demand expected for a student to correctly answer a test item. For example, an item or a task requiring students to predict a phenomenon based on data presented in a graph would generally be more complex than an item or task requiring students to simply describe the data presented in the graph.[23]

Comprehension score: A Growth Indicator. The Comprehension Score is computed from a single score estimation for each student using all reading and listening items.

Computer Adaptive Test/Testing (CAT): A technology-based assessment in which the next test item or set of items is selected by a computer algorithm, based on the student's performance on items administered earlier in the assessment. CAT algorithms allow the difficulty of the items presented on the assessment to be individually tailored based on the skill-level of each student being tested.[24]

Computer-Based Test (CBT): A technology-based assessment in which a computer or other computing device is used to deliver a test to the student.[25]

Conjunctive model: A conjunctive scoring model requires a minimum performance in all four domains when determining overall proficiency. Appling a rule that students must score at Level 4 or above in order to be proficient overall is an example of a conjunctive scoring model.

Construct: An abstract image or underlying theoretical concept that a test is designed to measure. Although the word construct is increasingly used in reference to what specific items are measuring, measurement traditionally defines it in terms of what a whole test measures.[26]

Constructed response item/Constructed Response (CR): A type of item on ELPA21 requiring a student response that is in a written or spoken, (e.g., short answer, essay, research report, oral presentation). The terms "open-ended" and "free-response" are sometimes used interchangeably with constructed-response.[27]

Content-specific: Specific to a given discipline, content area, domain, or subject area. (Within the literature and among researchers, the term "discipline-specific" is more commonly used.) CCSSO (2012) defines it as "the language used, orally or in writing, to communicate ideas, concepts, and information or to engage in activities in particular subject areas (e.g., science)" (p. 107). ELPA21 items are grounded within (but do not assess) three content areas: mathematics, English language arts, and science. [28]

---

[23]https://web.archive.org/web/20170626040152/http://parcconline.org/resources/parent-resources/glossary-of-terms
[24] CCSSO's Operational Best Practices 2013
[25] Ibid
[26] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja
[27] CCSSO's Operational Best Practices 2013
[28] The ELP Standards, page 211

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Content standards: Statements of the subject-specific knowledge and skills that teachers are expected to teach students at each grade level, indicating what students should know and be able to do.[29]

Content-Level Performance Level Descriptors (PLDs): Content-level PLDs indicate the knowledge, skills, and practices that students should be able to demonstrate at each performance level, in each content area (ELA/literacy and mathematics), and at each grade. Content- and grade level-specific PLDs are designed to inform test item development, the setting of performance level cut scores, as well as curricula and instruction at the local level.[30]

Criterion-Referenced Tests (CRT): A test that identifies a test taker's performance in relation to a predefined measurable level of mastery of skills (as compared to norm-referenced tests which compare a test taker's performance to the performance of other test takers; (see Achievement Test). ELPA21 is an example of a criterion-referenced test.[31]

Curriculum-based assessments: Assessments that measure the content covered in the curriculum, to use in an ongoing progress-monitoring and to inform adjustments in instruction, remediation, and student-accommodations. (See also *Instructionally-supportive tests*).[32]

Cut score: The point (or points) on a score scale that differentiates the interpretations made about those scoring above it from those scoring below it. Pass-fail, accepted-rejected, and proficient-not proficient are examples. Cut scores also are known as cutoff scores and performance standards.[33]

Designated features: These are additional features that can be provided to selected students during testing, as a means of increasing equitable access to the ELPA21. Educators familiar with their students' characteristics and needs should decide which students need these options. For the complete list of designated features available with the ELPA21 assessment system, refer to the ELPA21 Accessibility and Accommodations Manual.

Differentiated instruction: This is also referred to as "individualized" or "customized" instruction. The curriculum and instruction offer several different learning experiences within one lesson to meet students' varied needs or learning styles.

Disaggregated data: Information that is presented separately for various recognized student subgroups (e.g., gender, ethnicity, socio-economic status) or by other categories (e.g., grade, content area).[34]

---

[29] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja
[30] https://resources.newmeridiancorp.org/ela-test-design/
[31] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja
[32] ibid
[33] http://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf
[34] CCSSO's Operational Best Practices 2013

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Disaggregation: The collection and reporting of student achievement results by particular subgroups (e.g., students with disabilities, English learners, boys or girls, etc.).[35]

District assessment coordinator: The person responsible for managing the assessment program for the LEA or district. This person - the single point of contact and/or authority representing the district in the assessment system - coordinates with the service provider and/or the client regarding aspects of the assessment program, especially test administration within the district.[36]

Domain: The range of content covered by or targeted by a measure/test. ELPA21 measures student performance in the four (4) language domains recognized four language domains recognized by the US Federal Government: reading, writing, listening and speaking. [37]

Domain scores: A numeric three-digit scale score describes performance on the four domains of listening, reading, speaking and writing for the ELPA21 summative assessment. Each score is classified into one of five performance levels, where each level corresponds to a text descriptor stating what students in each level know and can do. The cut scores defining each level are documented in the ELPA21 Standard Setting Report and in *ELPA21 2016 Summative Assessment Scoring and Scaling Specifications* and the corresponding descriptors (ALDs) are available in the *ELPA21 2016 Achievement Level Descriptors by Domain and Grade.*

Domain test exemption: When a student is exempted from a domain test, no scale score is computed for that domain, and the performance level is assigned the letter code "E" or other similar administrative code (for "exempt").

Domain test not attempted: When a student does not attempt a domain test (but is not exempted from the domain), no scale score is computed for that domain, and the performance level is assigned the letter code "N" or other similar administrative code (for "not attempted"). Students with a "not attempted" for any domain may not be deemed proficient, regardless of performance level of the attempted domains. The missing domain shall be treated as the lowest possible score.

Elementary and Secondary Education Act (ESEA): Passed in 1965 as part of the "War on Poverty," the Elementary and Secondary Education Act emphasizes equal access to education and establishes high standards and accountability. The law authorizes federally funded education programs that are administered by the states. In 2002, Congress amended ESEA and reauthorized it as the No Child Left Behind).[38] In 2015 Congress amended ESEA and reauthorized it as the Every Student Succeeds Act.[39]

ELPA21 field test: The original ELPA21 Field Test took place February 2–March 31, 2015 and examined the ELPA21 test items and how they perform. Participation was voluntary.[40]

---

[35] http://www.edglossary.org/disaggregated-data/

[36] CCSSO's Operational Best Practices 2013

[37] ESSA Sec. 3113.

[38] http://www.k12.wa.us/esea/

[39] http://www.ed.gov/essa?src+rn

[40] Ibid

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Emerging: Students are Emerging when they have not yet attained a level of English language skill necessary to produce, interpret, and collaborate on grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile of Levels 1 and 2 in all four domains. Students scoring Emerging on ELPA21 are eligible for ongoing program support.

English Language Proficiency (ELP): The ability or capacity of individuals to use the English language in order to engage in the communication of ideas, knowledge, and information.[41]

English Language Proficiency Assessment for the 21st Century (ELPA21): ELPA21 is a test of English language proficiency. Derived from an innovative set of English language proficiency (ELP) Standards[42] developed during 2012–13, ELPA21 measures English learners' (ELs) ability to meet the language expectations required by grade-level English language arts, mathematics, and science content as specified by the Common Core State Standards[43] (CCSS), and the Next Generation Science Standards[44] (NGSS).

English language proficiency standards: Increasing the expectations for the academic content that students must master in grades K-12 requires a parallel increase in expectations for English language acquisition. The ELP Standards, to which the ELPA21 assessments align, describe these higher expectations by integrating language development with appropriate mathematics, English language arts, and science subject matter. As ELs learn and practice English in the classroom, they simultaneously interact with grade-level academic content. The ELP Standards describe higher expectations for ELs by integrating language development with appropriate mathematics, English language arts, and science practices by grade.

English Learner (EL): According to the USDOE, an English learner is an individual
  (A) Who is 3 to 21 years of age; and
  (B) Who is enrolled or preparing to enroll in an elementary or secondary school; and
  (C) (i) Who is a Native American or Alaska Native, or a native resident of the outlying areas; and
      (ii) Who comes from an environment where a language other than English has had a significant impact on the individual's level of English language proficiency; or
      (iii) Who is migratory, whose native language is a language other than English and who comes from an environment where a language other than English is dominant; and
  (D) Whose difficulties in speaking, reading, writing or understanding the English language may be sufficient to deny the individual –
      I.   The ability to meet the State's proficient level of achievement on State assessments described in Section 111 (b)(3)
      II.   The ability to successfully achieve in classrooms where the language of instruction is English; or

---

[41] See glossary definitions of "language proficiency" and "language competencies" from Council of Chief State School Officers (2012). F*ramework for English language proficiency development standards corresponding to the Common Core State Standards and the Next Generation Science Standards.* Washington, DC: Author. Retrieved from https://files.eric.ed.gov/fulltext/ED542756.pdf
[42] http://ccsso.org/resource-library/english-language-proficiency-elp-standards
[43] http://corestandards.org/
[44] http://www.nextgenscience.org/standards

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

III.   The opportunity to participate fully in society.[45]

ELPA21 applies this definition to defining ELs.

English learners with disabilities: Students who are defined as English learners (ELs) and have a 504 plan or an Individualized Education Program (IEP). Those with an IEP may be identified as having one or more categories of disability (autism, developmental delay, emotional disturbance, hearing impairment and deafness, intellectual disability, multiple disabilities, other health impairment, orthopedic impairment, specific learning disability, speech language impairment, traumatic brain injury, and visual impairment and blindness).[46]

Enhanced Assessment Grant (EAG): A grant offered to support state activities undertaken to improve the quality, validity, and reliability of state academic assessments beyond the requirements for such assessments described in section 1111(b)(3) of the Elementary and Secondary Education Act of 1965 (ESEA), as amended. Enhanced Assessment Grant funds may not be used for the routine printing, administration or scoring of existing state assessments. The grant funds may be used for the development of new assessment products or procedures, such as innovative test format, empirical analysis of variations in test format or procedures, or statistical models useful for combining data from multiple measures or for charting student progress over time.[47]

Errors of measurement: Errors in measurement refer to the amount of variation, or spread, in an examinee's test-score. A measurement error is the difference between an examinee's actual or obtained score and the unknowable "true" score. The Standard Error of Measurement (SEM) is a numerical value that is commonly used in interpreting and reporting individual test scores and score differences on tests.[48]

Ethical testing practices: Legitimate practices and procedures that must be maintained during the administration of a test. Unethical testing practices include inappropriate interactions between test administrators and students taking the test. They include, but are not limited to, allowing a student to answer fewer questions, offering additional information, coaching students during testing, editing student responses, or giving clues in any way.[49]

Every Student Succeeds Act (ESSA): The Every Student Succeeds Act (ESSA) was signed by President Obama on December 10, 2015. This bipartisan measure reauthorizes the 50-year-old Elementary and Secondary Education Act (ESEA), the nation's national education law and longstanding commitment to equal opportunity for all students.[50]

---

[45] Public Law 107-110. Title IX, Part A, Sec. 9101, (25)
[46] National Center on Educational Outcomes
[47] http://www.ed.gov/answers/
[48] See Harvill, L. M. (1991), Standard Error of Measurement. Educational Measurement: Issues and Practice, 10: 33–41
[49] National Center on Educational Outcomes
[50] retrieved from https://www.ed.gov/essa?src+rn, 8/24/17

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Evidence-Centered Design (ECD): An approach, followed by ELPA21, to constructing educational assessments that utilizes evidentiary reasons and arguments[51]. It requires developing a test from the start around the "inferences one wants to make, the observations one needs to ground them, the situations that will evoke those observations, and the chain of reasoning that connects them". (Messick, 1994)[52]

Exclusion: Federal law prohibits exclusion from ESEA testing and prohibits ELs from being excluded from English language proficiency testing based on a disability. However, in many states, parents/families may request that their student not participate in required federal testing for reasons such as medical conditions, religion, or personal beliefs.

Exemplar: Scored student work that evidences or exhibits an acceptable or particularly strong response for a particular rubric score point. Exemplars are usually used to train raters to score student performances, or to illustrate an acceptable level of performance to parents and the public.[53]

Exemption from testing: The act of releasing a student from a testing requirement to which other students are held.[54] An example of an exemption would be releasing a student who is hard of hearing from taking a listening test.

Fairness: Fairness of a test is attained when construct-irrelevant personal characteristics such as race, ethnicity, sex, or disability have no appreciable effect on test results or their interpretation.[55]

Field test: A test administration used during the test development process to check on the quality and appropriateness of test items, administration procedures, scoring, and/or reporting. Sometimes the field-test items are included as part of an operational test administration[56] (*Also see ELPA21 Field Test*).

Formative assessment: The use of assessments during the instructional process to monitor the progress of learning and the effectiveness of instruction so that adjustments can be made, as needed. This use is contrasted with the summative use of assessments.[57]

Gap analysis: 1) An assessment of the design of testing software to determine if all eventualities are addressed. 2) An assessment of missing pieces of a test. 3) An investigation of differences in achievement performance between two or more student subgroups, such as general education students and students with disabilities. 4) An analysis of available test items against test blueprints to determine future item-development needs.

---

[51] Mislevy, R.J., Steinberg, L.S., & Almond, R.G. (2002). On the roles of task model variables in assessment design. In S. Irvine & P. Kyllonen (Eds.), *Generating items for cognitive tests: Theory and practice* (pp. 97-128). Hillsdale, NJ: Erlbaum

[52] Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23

[53] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja

[54] ibid.

[55] http://nces.ed.gov/statprog/2002/glossary.asp

[56] http://www.ncme.org/resources-publications/books/testing-standards

[57] http://www.edglossary.org/formative-assessment/

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Grade appropriate: In the ELP Standards, this refers to level of content and text complexity aligned with the CCSS and NGSS requirements for a particular grade level or grade band. (See Appendix A of the CCSS for ELA & Language Standards and Defining the Core.)[58]

Grade band: The grade level or levels for which a particular test form or instance is designed. ELPA21 has test forms/instances for the following six grade bands: K, 1, 2-3, 4-5, 6-8, and 9-12.[59]

Grade level: The grade in which a student is enrolled or, for students not assigned a grade, the academic grade of same-age peers who are in a general education classroom.[60]

Grade-level achievement standards: Performance standards that reflect an understanding of the results of good teaching in the general curriculum in the enrolled grade. (See definition of *Performance Standards*.)[61]

Graduation rate: The percentage of students who graduated from high school as determined by a state-specific formula.

Growth indicators: ELPA21 provides two cross-domain measures of language ability, which states may use to measure and report growth. These Growth Indicators, 1) the Overall Score and 2) the Comprehension Score. They are not composites in the traditional sense, as weighted aggregates of separate domain scores, but is the total test scale score based on all items administered. Because these scores are computed differently than the domain scores, they are reported on a different reporting scale. Both are reported as a four-digit number.

High stakes testing: A test for which important consequences are attached to the results for students, teachers, schools, districts, states, and consortia. Consequences may include promotion, graduation, rewards, or sanctions.[62] ELPA21 tests are high stakes.

Home Language Survey (HLS): A survey administered to parents or guardians of every new student to collect information used to record the students whose families speak a language other than English at home. The survey also helps to identify the students who need to be assessed for English language proficiency.

Incomplete (but attempted) domain test: Once a domain test is considered "attempted" (started), any item on the form for which no response is provided (items that were omitted, skipped, or not reached) is assigned the minimum item score.

Individual Student Report (ISR): The Individual Student Report (ISR), provided by states to students and their parents or guardians, may describe:
1. A Determination of Proficiency

---

[58] The ELP Standards, page 213
[59] Dr. Phoebe Winter, Consultant
[60] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja
[61] Dr. Mary Seburn, Consultant
[62] http://www.edglossary.org/formative-assessment/

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

2. Performance level and descriptor for each domain score
3. Scale scores for each of the four domains
4. Error bands around each domain score (SEM, and probability of classification)
5. Explanatory text for concepts that may be easily understood or not commonly known
6. Interpretive guidance for interpreting each score
7. A clear explanation of all constructs measured by the test

Individualized Education Program (IEP): A written statement for each child with a disability that is developed, reviewed, and revised to reflect the decisions made by an interdisciplinary team, including the parent and the student when appropriate. An IEP for a student includes such components as instructional, placement, and assessment decisions.[63]

Individuals with Disabilities Education Act (IDEA): The Individuals with Disabilities Education Act (IDEA) is the nation's federal special education law that ensures public schools serve the educational needs of students with disabilities. IDEA requires that schools provide special education services to eligible students as outlined in a student's Individualized Education Program (IEP). IDEA also provides very specific requirements to guarantee a Free Appropriate Public Education (FAPE) for students with disabilities in the least restrictive environment (LRE). FAPE and LRE are the protected rights of every eligible child, in all fifty states and U.S. Territories.[64]

Instructional minutes: Refers to the amount of time the state requires teachers to spend providing instruction in each subject area.

Instructional scaffolding: Refers to the idea that specialized instructional supports need to be in place in order to best facilitate learning when students are first introduced to a new subject and/or for the purpose of ensuring that instruction provided in English is comprehensible to English Learners.

Instructionally-supportive tests: Tests that (1) supply clear and useable descriptions of what is to be assessed, and (2) provide instructionally informative results so that a student's mastery of each assessed curricular aim can be determined, and (3) guide further instruction and remediation.[65]

Integrated curriculum: A curriculum that combines content and skills from a variety of content areas in a unified way.

Interactive modality: The interactive modality emphasizes the need for ELs to meaningfully engage with their peers, instructors, and source materials during content area instruction. It is the collaborative use of receptive and productive modalities and refers to the learner as a speaker/listener and reader/writer. It requires two-way interactive communication where negotiation of meaning may be observed. The exchange will provide evidence of awareness of the socio-cultural aspects of communication as language proficiency develops. The interactive modality includes standards 2, 5, and 6.

---

[63] Ibid.
[64] http://www.parentcenterhub.org/idea/
[65] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Intercultural competence: The understanding and appreciation of cultural diversity when developing, administering, and interpreting assessments, and the capability to shift cultural perspective and adapt behavior to cultural similarities and differences along such cultural continua as language, ethnicity, race, disability, religion, place of origin, sexual orientation, gender, etc. [66]

Item: A question, exercise, task, or statement on a test for which the test taker is asked to select a response, create a response, or perform an activity that will be scored.[67]

Item acquisition: A process for the development of test items, typically including item prototypes, definitions of item development specifications, item review procedures, schedules, quantities, and quality acceptance criteria.

Item analysis: A part of the test development process in which the statistical properties of test items are carefully examined. Item analysis is sometimes used to determine the items that will be included in the final version of a test.

Item positioning: The location in the assessment in which an item is placed, relative to the beginning of the assessment, beginning of a section (i.e., after a break), or both. Item positioning is important in building fixed form tests as item positioning may impact student performance. (For example, a student may not do as well on some items if they are placed later in the test due to fatigue.)

Item specifications: A description of the major components to be included and considered in the development of test items including content, item types, cognitive complexity, rigor of the item, reading passage levels, and the use of graphics, tables, and charts.

Large-scale assessments: Tests that are administered to large groups of students within a district or state.[68]

Learning objectives: In education, learning objectives are brief statements that describe what students will be expected to learn by the end of school year, course, unit, lesson, project, or class period. In many cases, learning objectives are the interim academic goals that teachers establish for students who are working toward meeting more comprehensive learning standards.[69]

Linking: The process of placing scores from two test forms onto a common score scale. When the test forms are parallel, the specific process is called equating.[70]

Longitudinal methods: The comparison of assessment results or other outcomes and behaviors of the same groups of students collected at two or more points over an extended period of time.[71]

---

[66] National Center on Educational Outcomes
[67] http://www.ncme.org/resources-publications/books/testing-standards
[68] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja
[69] http://edglossary.org/learning-objectives/
[70] http://www.ncme.org/resources-publications/books/testing-standards
[71] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Machine scoring: An automated system for scoring test takers' responses to items (e.g., selected response, technology enhanced, drag-and-drop), that can be scored as correct or incorrect.

Matrix sampling: A method of test administration in which different subsets of students (often assigned at random) are administered different subsets of items. Matrix sampling is often used when there is interest in a large number of test items but constraints on the number of items that can reasonably be administered to any one student.

Measurement model: The Measurement Model is that part of the Evidence Model that explains how the scoring should be updated given the observed features of the response.[72]

Measuring progress: ELPA21 scores can be used to monitor progress by ELs towards English language proficiency and for describing individual and group strengths by domain and over time. Reliably measuring progress over time meets multiple state needs such as informing student placement and program reclassification, determining instructional needs of ELs and the support needs of EL teachers, evaluating program effectiveness for subgroups of students, and adjusting educational programming and resources as needed.

Metadata: A term used to represent all information and attributes associated with a test item other than the item itself (e.g., item type, font size, formatting, statistics, value of weight of the item, correct response key, content standard measured).[73]

Modalities: The ELP Standards describe three modalities: receptive, productive, and interactive. They are the characteristics of the "channels," or modes of communication through which language is used.

Multidimensional: A test or subtest is multidimensional if it measures more than one construct. In English language proficiency assessment, multiple constructs can include domain (reading, writing, listening, and speaking), mode (receptive, productive, interactive), and academic subject (math, science, ELA). Multidimensional assessments require special analytic methods to correctly measure and score each construct.

Multiple measures: Measurement of student or school performance using multiple methods or instruments for evaluating knowledge or skills. For students, these might include teacher observations, performance assessments, or portfolios. For schools, these might include dropout rates, absenteeism, college attendance, or documented behavior problems.[74]

National Assessment of Educational Progress (NAEP): A national test that is given to specific grade levels in specific subjects. A small sample of students who are representative of the state is tested.

---

[72] Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2004). A brief introduction to evidence centered design. Educational Testing Service (ETS). https://doi.org/10.1002/j.2333-8504.2003.tb01908.x
[73] CCSSO's Operational Best Practices 2013
[74] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

NAEP test scores can be used to compare student performance across states and to examine student performance across the country on a common measure.

Natural cue: An environmental cue that allows the student to perform a task without a direct or indirect cue from an instructor.[75]

Norm-Referenced Tests (NRT): A standardized test designed, validated, and implemented to rank a student's performance by comparing that performance to the performance of the student's peers, or other reference group.[76]

Objective item: Also referred to as selected response items, are an item type for which the possible correct responses for the student to select from are limited and predefined. Objective items have an answer key that indicates the correct and incorrect responses.[77]

On-demand score report: Predefined score report of student, group or institutional performance, generated upon request and delivered electronically.[78]

Open-ended item: See Constructed-response item.

Operational assessment: An assessment that is developed and administered at a specified time and whose scoring results are used for the purpose of meeting local, state, and/or federal requirements.[79]

Operational item: A test item used to contribute to a student's test score.

Opportunity to Learn (OTL): OTL refers to the equitable provision or distribution of conditions and resources (e.g., curricula, learning materials, facilities, equipment, and teachers) within a school or classroom to provide balanced opportunities for all students to learn, regardless of disability or other student characteristics.[80]

Out-of-level testing: Administration of a test at a level above or below a student's present placement grade level to enable the student to be assessed at the level of instruction rather than the level of enrollment. Testing students below their grade of enrollment is not in compliance with the 1% regulation of December 9, 2003 unless it is used as an alternate assessment based on alternate achievement standards for students with the most significant cognitive disabilities, and has been subjected to the same requirements as those assessments (e.g., rigorous standards setting).[81]

Overall score: A Growth indicator. The Overall Score is computed from a single score estimation for each student including items from all four domains.

---

[75] Ibid

[76] Ibid

[77] CCSSO's Operational Best Practices 2013

[78] Ibid

[79] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja

[80] Ibid

[81] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Paper-based/Paper-and-pencil: A form of ELPA21 delivered in a printed hard copy form, rather than in a digital form.[82]

Participation criteria: Criteria which define how ELs and ELs with disabilities participate in assessments, and criteria that determines which assessments these students take. Students participate in the regular assessment, English language proficiency assessment, or alternate assessment guided by each state's participation criteria for inclusion. These criteria define the ways in which students may participate in the assessment system, and the criteria for determining which assessment a given student will take.[83]

Participation rate: The percentage of students within a particular group or subgroup who took a test as compared to the total number of students in the group or subgroup who could have taken the test.[84]

Partnership for Assessment of Readiness for College and Careers (PARCC): The Partnership for Assessment of Readiness for College and Careers (PARCC) was a group of states working together to develop a set of assessments that measured whether students were on track to be successful in college and their careers. The computer-based, K-12 assessments in Mathematics and English Language Arts/Literacy gave teachers, schools, students and parents better information regarding student learning, indicated likelihood of success after high school and provided tools to help teachers customize learning to meet student needs. In 2017, the PARCC assessment consortium transitioned to New Meridian Corp., a nonprofit organization expanding assessments to various content areas and working to create flexible and dynamic assessments.[85]

Performance assessment: A form of testing that requires a student to perform a task, (e.g., write an essay, design or conduct a laboratory experiment, or maintain a portfolio), rather than select an answer from a pre-made list (e.g., multiple choice items).[86]

Performance levels: Performance levels are broad categories that describe the results of an assessment, and ELPA21 has five (Beginning, Early Intermediate, Intermediate, Early Advanced, and Advanced). These levels describe the stages of English language development through which ELs are expected to progress as they gain proficiency.

Performance standards: A description of the minimum performance required to judge a test taker's score to be "good enough". The description is used in a standard setting procedure to determine the cut score required to differentiate performance levels, such as "pass" rather than "fail", eligible for certification rather than not eligible, or proficient rather than not proficient.[87] Also known as "cut scores."

---

[82] Council of Chief State School Officers (2013). Operational Best Practices for Statewide Large-Scale Assessment Programs.
[83] NCEO Brief 3: http://nceo.info/Resources/publications/OnlinePubs/briefs/brief03/brief03.html
[84] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja
[85] www.parcconline.org
[86] https://learningpolicyinstitute.org/product/cpac-performance-assessments-support-student-learning-brief
[87] http://www.ncme.org/resources-publications/books/testing-standards

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Performance target: Proficiency as measured by ELPA21 requires meeting a combination of expectations across all four domains. This expectation represents the knowledge, skills and abilities that are required in each domain to interact with and engage in grade-level content instruction and is referred to as the "performance target."

Performance Level Descriptors (PLDs): A description of what performance at each level of the test should and does look like. PLDs describe the range of performance the test measures and are generally created for the test overall, and by grade and domain (such as Science, Math, and ELA) for each level of performance. PLDs are also known as Achievement Level Descriptors. In score reporting, they describe what scores mean and communicate what students scoring at each level know and are able to do.[88]

Pilot test: A stand-alone administration of test items, tools, or a system, to evaluate how particular items function prior to a field test and operational use. The pilot test generally occurs with a sample of students that matches the purpose of the pilot.[89]

Platform: Refers to the method that will be used to deliver the presentation materials to the examinees. Platform is broadly defined to include human examiners, computer, paper and pencil, etc., and is specifically defined as the environment in which testing software and/or applications are designed to run.[90]

Policy Definitions (PDs): A type of performance descriptor. Describe the rigor of, and ELPA21's vision for, English language proficiency and its impact on policy, consistent across grade.

Portfolio assessment: A portfolio is a collection of student-generated or student-focused work that provides the basis for demonstrating the student's mastery of a range of skills, performance level, or improvement in these skills over time. A portfolio becomes a portfolio assessment when (1) the assessment purpose is defined; (2) criteria or methods are made clear for determining what is put into the portfolio, by whom, and when; and (3) criteria for assessing either the collection or individual pieces of work are identified and used to make judgments about performance. The portfolio evidence may include student work samples, photographs, videotapes, interviews, anecdotal records, interviews, and observations.[91]

Productive modality: The productive modality places the learner as speaker and writer for a 'distant' audience, one with whom interaction is limited or not possible. It is a planned or formalized speech act or written document, and the learner has an opportunity to draft, get feedback, and revise, before publication or broadcast. The productive modality requires spoken and written language skills (the speaking and writing domains) and includes standards 3, 4, and 7.

---

[88] Dr. Mary Seburn, Consultant
[89] CCSSO's Operational Best Practices 2013
[90] Dr. Mary Seburn, Consultant
[91] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Proficiency: Mastery or ability to use the English language at the level required by rigorous grade-level content standards without requiring EL Program support.

Proficiency determination: Using the profiles, different combinations of skills and abilities across the domains are deemed as "Proficient", "Progressing" or "Emerging." The Proficiency Determination (often referred to by states as the Overall Proficiency Determination) identifies ELs whose language proficiency enable full participation in grade-level academic contexts.

Progressing: Students are Progressing when, with support, they approach a level of English language skill necessary to produce, interpret, and collaborate, on grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile with one or more domain scores above Level 2 that does not meet the requirements to be Proficient. Students scoring Progressing on ELPA21 are eligible for ongoing program support.

Race to the Top (RTTT): A competitive federal grant program run by the U.S. Department of Education that began in 2009 and provides a total of $4 billion in one-time grants to a handful of states that have created conditions for bold, comprehensive action in four reform areas described in the American Recovery and Reinvestment Act (ARRA).[92]

Rangefinding: When conducting the scoring of constructed response items/assessments, a process used to identify a range of student responses for each identified score point or level, articulating performance as defined by a scoring rubric.[93]

Rating scale: A rating scale is used to score a student's response by assigning values to that response. These ratings are based on a scoring tool (See *Rubric*) to ensure consistent scoring across students and raters. Rating scales include numerical scales or descriptive scales.[94]

Raw score: The score obtained by a test taker reflecting the number of items correctly answered, or the number of points awarded during scoring, for the responses given. Raw scores often are converted to some kind of derived score, such as percentile rank, scale score, or performance level, for interpretive purposes.[95]

Readability: Readability refers to the extent to which a reader at a particular level can be expected to be able to comprehend text. There are different formulas that are used to estimate the ease with which a document or passage can be read. Readability helps to assess the suitability of reading material for students at particular ages or grade levels. Readability formulas consider the organization of text, syntactic complexity of sentences, use of abstractions, density of concepts, sequence and organization of ideas, page format, sentence length, paragraph length, variety of punctuation, student background knowledge and/or interest, and use of illustrations or graphics in determining the appropriate level of difficulty of instructional or assessment materials. Readability has been used

---

[92] http://www.fcc.gov/general/american-recovery-and-reinvestment-act-2009
[93] CCSSO's Operational Best Practices 2013
[94] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja
[95] http://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

to assess the reading passages in tests to ensure age/grade appropriateness and to increase confidence that the level of difficulty does not interfere with the task requirement.[96]

Real-world application: The opportunity for a student to exhibit a behavior or complete a task that the student would normally be expected to perform outside of the school environment and in activities not associated with the educational process.[97]

Receptive modality: The receptive modality refers to the learner as a reader and listener/viewer working with 'text' whose author or deliverer is not present or accessible. It presumes authentic written or oral documents where language input is meaningful and content laden. The receptive modality requires skills necessary for interpreting and comprehending spoken or written messages and includes standards 1 and 8.

Reclassification: ELPA21 scores can be used to determine proficiency relative to grade appropriate performance standards for reclassification purposes. Once proficient, ELs will have acquired the content-specific English language practices that enable them to produce, interpret, collaborate on, and succeed in content-related and grade-appropriate academic tasks.

Reference group: The group of students with which an individual or other group is compared in a study of differential item functioning; often this is the largest group of students. This may also refer to the group of students to which scores on a norm-referenced test are compared.[98]

Reliability: The degree to which 1) the scores of every individual are consistent over repeated applications of a measurement procedure and hence are dependable and repeatable; 2) the degree to which scores are free of errors of measurement. Reliability is usually expressed in the form of a reliability coefficient or as the standard error of measurement derived from it. The higher the reliability coefficient the better, because this means there are smaller random errors in the scores.[99]

Reporting performance level descriptors: A type of performance descriptor. Derived from the final PLDs, they describe the appropriate inferences that may be made about the students who score in each performance level.

Response option distribution: A summary of the number of students selecting each of the response options on a multiple-choice item (e.g., where "A, B, C, etc." are used to label response choices, how many correct answers are "As", how many "Bs," etc.) The distribution is calculated to determine that there is an acceptable balance of correct responses among the options.[100]

---

[96] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja
[97] Ibid
[98] https://nces.ed.gov/nationsreportcard/glossary.aspx?nav=y
[99] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja
[100] CCSSO's Operational Best Practices 2013

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Response requirements: The type, kind, or method of action required of a student to answer a question or test item. The response may include, but is not limited to, reading, writing, speaking, creating, and drawing.[101]

Roster score report: States will provide roster reports to school administrators and educators summarizing their students' performance. Member states indicated ELs tend not to be assigned to a single classroom, but to multiple grade-level classrooms and suggested that summary reports may be most useful at the grade- rather than classroom-level. As a result, states may choose to group students in these reports using the categorization scheme that is most useful to each state.

Rubric: A scoring tool based on a set of criteria used to evaluate a student's ELPA21 test performance. The student's response can be compared to the descriptions contained in the rubric to determine the appropriate score to assign to the response. The criteria contain a description of the requirements for varying degrees of success in responding to the question or performing the task. Rubrics may be diagnostic, analytic (i.e., providing ratings of multiple criteria), or holistic (i.e., describing a single, global trait).[102]

Sample: A subset of a population whose characteristics are studied to gain information about the entire population.[103]

Sample type: A designation given to a sample to indicate which administration rules were used: (e.g., either those that did not allow accommodations or those that did allow accommodations).[104]

Sampling error: The error in survey estimates that occurs because only a sample of the population is observed. Measured by sampling standard error.[105]

Sampling frame: The list of sampling units from which the sample is selected.[106]

Sampling variability: The variability in survey estimates that occurs because only a sample of the population is observed. Measured by standard error.[107]

Sampling weight: A numeric value calculated on the basis of each respondent's characteristics (gender, ethnicity, year of birth, sample type, and location).[108]

Scale score: A kind of score to which a raw score has been converted to a numeric scale for ease of interpretation.[109]

---

[101] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja
[102] http://www.edglossary.org/
[103] https://nces.ed.gov/nationsreportcard/glossary.aspx?nav=y
[104] Ibid
[105] Ibid
[106] Ibid
[107] Ibid
[108] https://www.nlsinfo.org/content/getting-started/intro-to-the-nls/glossary-nls-terms/glossary-nls-terms-altogether/glossary
[109] http://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Scaling: The process of assigning a scale score based on the pattern of responses.[110]

Scientifically-based research: Research that involves the application of rigorous, systemic, and objective procedures to obtain reliable and valid knowledge relevant to educational activities and programs.

Score scale: A scale used to describe what students know and can do.[111]

Scoring rubrics: A guide used to score a response to a constructed-response item. Also referred to as *Scoring Guide.*

Screener: An assessment intended to determine whether a student is eligible or ineligible for a service or program. The ELPA21 screener is designed to assist in deciding whether a student is eligible for EL services.[112]

Section 504: Section 504 of the Rehabilitation Act of 1973 requires public schools to provide accommodations to students with disabilities even if they do not qualify for special education services under Individuals with Disabilities Education Act (IDEA). The definition of a disability under Section 504 is much broader than the definition under IDEA. All IDEA students are also covered by Section 504, but not all Section 504 students are eligible for services under IDEA *(see also Students with disabilities*). Section 504 states: No otherwise qualified individual with a disability in the United States shall, solely by reason of her or his disability, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance. (29 U.S.C. Section 794)

Section 508: Section 508 is part of the Rehabilitation Act of 1973 which requires that electronic and information technology developed, procured, maintained, or used by the Federal government be accessible to people with disabilities. Section 508 requires that individuals receive information by alternative means if an agency is not 508 compliant.

Secure forms of assessments: Tests that will be used repeatedly with different groups of students and must be safeguarded so all students have equal exposure to the test materials and equal opportunities for success.[113]

Secure materials: Test materials (e.g., an assessment, administration direction, teacher guidelines) that must be protected from release, either, prior to test administration or because they will be reused in the future. Holding test materials secure is necessary to ensure the integrity of an assessment.[114]

---

[110] Ibid
[111] https://nces.ed.gov/nationsreportcard/glossary.aspx?nav=y
[112] Dr. Phoebe Winter, Consultant
[113] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja
[114] CCSSO's Operational Best Practices 2013

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Selected Response (SR) items: More commonly known as multiple choice items. On ELPA21, these are items that allow the student to choose a response from a group of two or more provided responses.

Smarter Balanced Assessment Consortium: The Smarter Balanced Assessment Consortium (Smarter Balanced) is one of two multistate-led consortia working to develop next-generation assessments aligned with the Common Core State Standards that accurately measure student progress toward college- and career-readiness.[115]

Special population: A term used to identify a disadvantaged group of students. Special populations may include such groups as students with disabilities, low-performing students, students from low-income families, etc. Special populations often require accommodations for physical, mental, or emotional differences.[116]

Stakeholders: A group of individuals perceived to be vested in a particular decision. These individuals might include parents, educators, students, and community leaders.[117]

Standard Error of Measurement (SEM): A test score on a standardized assessment represents an estimate of a student's unknown 'true' score. If the student were to take the same test many times (hypothetically), the standard deviation of these scores is defined as the standard error of measurement (SEM) and is expressed in the same scale units as the reported test score. The SEM is used to represent the precision of the observed score. Plus and minus one SEM gives the expected range in which the student's score would fall 68 percent of the time.

Standard setting: The process of identifying the scores (cut scores) on a score scale that define the starting and ending points of the performance levels used for reporting test performance. For example, the process of standard setting is used to determine the lowest score that can categorize performance as **"proficient"**.[118]

Standardization: An established procedure that ensures a test is administered with the same directions, under the same conditions and is scored in the same manner for all students to ensure the comparability of scores. Standardization allows reliable and valid comparison to be made among students taking the test. The two major types of standardized tests are norm-referenced and criterion-referenced.[119]

Standardized Achievement Test (SAT): Also known as the SAT Reasoning Test (formerly called Scholastic Aptitude Test), this test is widely used as a college entrance examination. Scores can be compared to state and national averages of seniors graduating from any public or private school.

---

[115] http://smarterbalanced.org/
[116] National Center on Educational Outcomes
[117] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja
[118] http://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf
[119] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Standards-based assessments: Assessments constructed to measure how well students have mastered specific content standards or skills.[120]

Stratified sampling design: A procedure for selecting the individuals who are to be included in a study that involves first dividing the population into groups based on some characteristic or set of characteristics, then selecting (usually at random) individuals from each group.

Student demographics: Information on student characteristics, such as grade level/band, age, gender, ethnicity, primary language, disability, time spent in the United States, and other categories.[121]

Students with disabilities: Students with disabilities include students who have 504 accommodation plans and students who have Individualized Education Programs (IEPs). Those with an IEP may be identified as having one or more categories of disability (autism, deaf blind, developmental delay, emotional disturbance, hearing impairment and deafness, intellectual disability, multiple disabilities, other health impairment, orthopedic impairment, specific learning disability, speech language impairment, traumatic brain injury, and visual impairment and blindness).[122]

Students with significant cognitive disabilities: Students who are within one of the existing categories of disability under IDEA (autism, deaf-blindness, hearing impairment, mental retardation, orthopedic impairment, deafness, emotional disturbance, multiple disability, traumatic brain injury, visual impairment, learning disability, speech and language impairment, other health impaired) whose cognitive impairments may prevent them from attaining grade-level achievement standards, even with the very best instruction.[123]

Subgroup: A well-defined group of students. For example, ESEA identifies the following specific subgroups that must achieve adequate yearly progress: students of racial or ethnic minority, students with disabilities, limited-English-proficient, English learners, economically disadvantaged students, etc. [124]

Summative assessment: A comprehensive assessment designed to provide information about the level of student, school, or program success at a defined point in time, typically end of year. Summative tests are administered after the conclusion of instruction.[125]

Sustainability: ELPA21 sustainability is a process to define and develop post grant options on how collaborative states will procure, administer, maintain and improve the ELPA21 assessment following the period of federal grant funding. Sustainability is defined by five domains including: transition planning and funding opportunities; identification of necessary sustainability elements; staffing and governance; continued improvements; and fiscal sustainability.[126]

---

[120] Ibid

[121] National Center on Educational Outcomes

[122] Ibid

[123] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja

[124] Ibid

[125] National Center on Educational Outcomes

[126] Dr. Susan Inman, Consultant

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Target student descriptors: A type of performance descriptor. Also called *target* or *standard setting* performance level descriptors, or *target student descriptions*, typically derived from the range PLDs to describe the *minimum* policy and content expectations for each performance level. They describe the minimal skills that barely proficient students in each grade band should possess.

Task: A task is a unit of work requested from an examinee during the course of an assessment.[127]

Technical adequacy: Empirical and procedural evidence that supports the specific instructional and policy decisions being made from the use of an assessment.[128]

Technical document: A document that provides the technical and psychometric information on an assessment. (Sometimes referred to as a Technical Report or Technical Manual).[129]

Technology-based: An assessment in digital form that is delivered through the use of technology, rather than administered on a paper-based form.[130]

Technology-Enhanced Items (TEIs): On ELPA21, TEIs are items administered on a computer that take advantage of the computer-based environment to present situations and capture responses in ways that are not possible on a paper-based test.[131]

Test: A measuring device or procedure. Educational tests are typically composed of questions or tasks designed to elicit predetermined behavioral responses or to measure specific academic content standards.[132]

Test design: Specifications (e.g., assessment type, desired layouts, use of supporting materials) defining the nature of a test.[133]

Test forms: Versions of ELPA21 that are considered interchangeable in that they measure the same constructs, are intended for the same purposes, and are administered using the same directions.[134]

Test map: Identification of the order and the manner in which test items will be positioned in a form, whether paper or digital.[135]

---

[127] Mislevy, R. J., Almond, R. G., & Lukas, J. (2004). A brief introduction to evidence-centered design (CSE Technical Report No. 632). The National Center for Research on Evaluation, Standards, Student Testing (CRESST). Available from http://cresst.org/publications/cresst-publication-3003 (Also ETS Research Report RR-03-32.)

[128] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja

[129] Ibid

[130] CCSSO's Operational Best Practices 2013

[131] Ibid

[132] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja

[133] CCSSO's Operational Best Practices 2013

[134] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja

[135] CCSSO's Operational Best Practices 2013

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Test presentation: The method, manner, or structure in which test items or assessments are administered to the student.[136]

Test security: Established procedures to ensure current or future confidentiality, fidelity, and integrity of a test whereby public access is limited and strictly monitored, with clearly outlined consequences for breaches in test security.[137]

Test specifications: A description of the major components to be included and considered in the development of a test including content, item types, cognitive complexity, rigor of items, reading passage levels, and the use of graphics, tables, and charts. This includes written details prepared early in the test development process to describe many of the characteristics of the resulting test. Generally included in such a document are: the content to be covered, the level of cognitive functioning to be displayed with the content, the proportional weighting of the various content areas, the kinds of items to be used, the number of items of each type to be used, and how much time will be permitted for testing. Brief versions of the specifications are sometimes referred to as test plans or test blueprints.[138]

Testing irregularity: Conduct by either a student or an administrator during ELPA21 testing that is not part of the standardized procedures established for the handling of secure test materials, and/or the established standardized test administration protocols.[139] Test irregularities may invalidate test scores that were obtained during the irregularity.

Title I: The Title I program originated as Title I of the Elementary and Secondary Act of 1965. Its primary purpose was to ensure that all children were given the opportunity to be provided with a high-quality education. Title I is the largest federally funded education program for elementary and secondary schools. Title I is also designed to focus on special needs populations and to reduce the gap between advantaged and disadvantaged students.

States that have Title I schools have several requirements to keep the funding. Schools must have a child poverty rate of at least 40% to operate a school wide Title I program. A school wide Title I program can provide benefits to all students and is not just limited to those students who are considered to be economically disadvantaged.[140]

Title III: Part A of Title III is officially known as the *English Language Acquisition, Language Enhancement, and Academic Achievement Act*. Section 3102 lists five purposes of the law. The overarching purpose is to ensure that English learners, including immigrant children and youths, attain English proficiency and meet the same challenging academic content and achievement standards that other students are expected to meet.

Local Education Agencies (LEAs) must use Title III funds to implement language instruction

---

[136] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja
[137] Ibid
[138] http://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf
[139] Council of Chief State School Officers (2013). Operational Best Practices for Statewide Large-Scale Assessment Programs.
[140] http://teaching.about.com/od/s-zteachingvocabulary/g/Title-I.htm

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

educational programs designed to help LEP students achieve standards. The state educational agency (SEA), LEAs, and schools are accountable for increasing the English proficiency and core academic content knowledge of LEP students.

With the authorization of the Every Student Succeeds Act (ESSA) in December 2015, Title III was moved under Title I.

Title III accountability: Title III of the *Elementary and Secondary Education Act (ESEA)* provides formula-based grants to states to help support the education needs of students identified as English Learners (ELs). States distribute these Title III funds to local school districts and consortia of districts through subgrants. As a condition of funding, Title III requires states to design and implement an accountability program under which districts and other subgrantees are expected to meet targets related to EL growth in both language proficiency and academic achievement.

Title III's accountability requirements dovetail with the broader academic accountability requirements established under Title I. Under Title I of *ESEA*, states are required to establish annual assessments to measure student academic achievement in reading and math. EL students participate in this annual testing, and their performance is reported in the "all students" category and in a special EL subgroup.

Under Title III of *ESEA*, states must establish standards and assessments related to English language proficiency (ELP) standards and assessments.

Twinned item: An item designed to replace another item, measuring the same content at the same level of complexity and difficulty. Twinned items are usually developed to provide a comparable item in another format when the original item cannot be directly translated to that format, for example, a technology-enhanced item that needs to be rendered on paper or in braille. (In some cases, for example with some multi-point technology-enhanced items, more than one replacement item is needed to create a twin in another format.)[141]

Universal Design (UD) for assessment: Describes a framework for curriculum design, instructional processes and tests that provides all students with equal opportunities to learn and demonstrate their knowledge and skills. The purpose is to develop tests that are accessible to as many students as possible. Universal design builds flexibility into curricula and tests at the point of development, which enhances a teacher's ability to make adjustments for different learners during classroom instruction. Using these principles, test developers consider the full range of students being tested and develop items, tasks and prompts that measure learning for the greatest number of students without the need for accommodation if possible. Assessments developed using this framework have been designed from the beginning to ensure accessibility by all students regardless of ability or disability. At this time, research is being conducted to better delineate how to build universally designed assessments.[142]

---

[141] Dr. Phoebe Winter, Consultant and Dr. Martha Thurlow, NCEO
[142] https://ucla.app.box.com/file/1425249515746?s=3vp8s6clhetb6va0fhu0vqhowe4ny7ja

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Universal features: These features are available to all students during testing, as a means of increasing equitable access to the ELPA21. Selection or deselection of these features is usually based on student preference. For the complete list of universal features available with the ELPA21 assessment system, refer to the ELPA21 Accessibility and Accommodations Manual.

Validation: The process of gathering evidence to (a) support particular meanings that the user would like to attribute to scores from a test and (b) demonstrate that it is appropriate to use the scores from that test in the way(s) the user has chosen. Though the burden generally is on the user of test scores to provide the evidence, often the test developer furnishes information about test purpose, test development procedures, and appropriate test administration and scoring procedures.[143]

Validity: The extent to which a test or set of operations measures what it is supposed to measure. The degree to which the evidence obtained through validation supports the score interpretations and uses to be made of the scores from a certain test administered to a certain person or group on a specific occasion. Sometimes the evidence shows why competing interpretations or uses are inappropriate, or less appropriate, than the proposed ones.[144]

Vertical alignment: Planning curriculum across the grade levels, from Kindergarten through high school, building upon instruction based upon standards.

Vertical scale: A single scale that allows for tracking student growth and progress across grades and over time.[145]

---

[143] http://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf
[144] Ibid
[145] Dr. Mary Seburn, Consultant

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

ELPA21 Acronyms Glossary

504 - Section 504 of the *Rehabilitation Act of 1973*

508 - Section 508 of the *Rehabilitation Act of 1973*

AA - Alternate Assessment

AA-AAS - Alternate Assessments Based on Alternate Achievement Standards

AAC - Alternative and Augmentative Communication

AASA - American Association of School Administrators

ADA - *Americans with Disabilities Act of 1990*

AERA - American Educational Research Association

ALD - Achievement Level Descriptor

ALI - Achievement Level Indicator

AMAO - Annual Measurable Achievement Objectives

ANOVA - Analysis of Variance

APIP - Accessible Portable Item Profile

ASDS - Assessment and Standards Development Services

ASES - Assessing Special Education Students

ASG - Assessment Solutions Group

AYP - Adequate Yearly Progress

BOE - Body of Evidence

CAS - Comprehensive Assessment Systems for Title I

CAST - Center for Applied Special Technology

CAT - Computer Adaptive Testing

CBT - Computer-Based Test

CC - Consortium Council

CCR - College and Career Readiness

CCSS - Common Core State Standards

CCSSO - Council of Chief State School Officers

CFR - Code of Federal Regulations

CR - Constructed-Response

CRESST - Center for Research, Evaluation, Standards and Student Testing

CROS - Constructed-Response with Objective Scoring

CRSS - Constructed-Response with Subjective Scoring

CRT - Criterion-Referenced Test

CSAI - Center for Standards and Assessment Implementation

DLM - Dynamic Learning Maps

EAG - Education Action Group

EAG - Enhanced Assessment Grant

EB - Executive Board

ECD - Evidence-Centered Design

ECEA - Early Childhood Education Assessment Consortium

EDGAR - Education Department General Administrative Regulations

EHA - *Education of the Handicapped Act*

EL – English Learner

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

ELA - English language arts

ELDA - English Language Development Assessment

ELL - English language learner

ELP - English language proficiency

ELPA21 - English Language Proficiency Assessment for the 21st Century

EPK - Electronic Press Kit

EPRRI - Education Policy Reform Research Institute

ERIC - Educational Resources Information Center

ESEA - Elementary and Secondary Education Act of 1965

ESL - English as Second Language

FAPE - Free and Appropriate Public Education

FERPA - Family Educational Rights and Privacy Act of 1974

FTE - Full Time Equivalent

GEPA - *General Educational Provisions Act*

GSEG - General Supervision Enhancement Grant

HEAP - Health Education Assessment Project

HLS - Home Language Survey

IASA - *Improving America's Schools Act*

ICCS - Implementing the Common Core Standards

IDEA - *Individuals with Disabilities Education Act of 2004*

IEP - Individualized Education Program

IES - Institute of Education Science

IHE - Institutions on Higher Education

IRT - Item Response Theory

ISR - Individual Student Report

LEA – Local Educational Agency

LRE - Least Restrictive Environment

MCREL - Mid-Continent Regional Educational Laboratory

m-IRT - multi-dimensional Item Response Theory

MOU - Memorandum of Understanding

MSRRC - Mid-South Regional Resource Center

MTSS – Multi-Tiered System of Supports

NAAC - National Alternate Assessment Center

NAEP - National Assessment of Educational Progress

NAESP - National Association of Elementary School Principals

NAGB - National Assessment Governing Board

NASBE - National Association of State Boards of Education

NASDSE - National Association of State Directors of Special Education

NASP - National Association for School Psychologists

NASSP - National Association for Secondary School Principals

NCAC - National Center on Accessing General Curriculum

NCEO - National Center on Educational Outcomes

NCES - National Center for Education Statistics

NCLB - *No Child Left Behind Act of 2001*

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

NCREL - North Central Regional Educational Laboratory

NCRRC - North Central Regional Resource Center

NCSC - National Center and State Collaborative

NERRC - Northeast Regional Resource Center

NGSS - Next Generation Science Standards

NRT - Norm-Referenced Test

NSBA - National School Boards Association

OCR - Office for Civil Rights

OERI - Office of Educational Research and Improvement

OSEP - Office of Special Education Programs

OSERS - Office of Special Education and Rehabilitation Services

OTL - Opportunity to Learn

PDM - Professional Development Modules

PLD - Performance Level Descriptor

RFP - Requests for Proposal

RRC - Regional Resource Center

RTTT - Race-to-the-Top

SAT - Standardized Achievement Test

SCASS - State Collaborative on Assessment and Student Standards

SEA - State Education Agency

SEC - Surveys of the Enacted Curriculum

SERRC - Southeast Regional Resource Center

Smarter Balanced - Smarter Balanced Assessment Consortium

SR - Select-response

SWD - Students with Disabilities or Student with a Disability

TAC - Technical Advisory Council/Committee

TDS – Test Delivery System

TE - Technology Enhanced

TECR – Technology Enhanced Constructed Response (item type)

TEI – Technology Enhanced Item

TESOL - Teachers of English to Speakers of Other Languages

TIDE – Test Information and Distribution Engine

TILSA - Technical Issues in Large-Scale Assessment

TIMSS - Third International Mathematics and Science Study

UAT – User Acceptance Testing

UDA - Universal Design for Assessment

UDL - Universal Design for Learning

US DOE - U.S. Department of Education

WIDA - World-class Instructional Design and Assessment English Language Development

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

Document Change History

Date	Version	Changes Made by	Description of Change
10/8/14	1.0	N/A	Release Version
10/9/14	1.1	Lauren Lynch	Revised typographic errors, Updated definition of "Title I", page 25.
6/10/15	1.2	Lauren Lynch	Revised definitions per Consortium Council review
10/01 /17		Margaret Ho, Michelle McCoy, Jamie Lam	Updated and finalized information from contributions from state members on the Member Services & Outreach Committee, and the Assessment Design Team.
11/17/23		Gabriela Venegas	Branding and copyright updates, review of links toward finalization.
1/30/24		Michelle McCoy, Brooke Jambor, Eric Shekoufeh	Update expired links, branding, and copyright as needed.

English Language Proficiency Assessment for the 21st Century
National Center for Research, Evaluation, Standards, and Student Testing
University of California, Los Angeles

English Language Proficiency Assessment for the 21st Century

Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Bldg., Box 951522
Los Angeles, CA 90095-1522

www.elpa21.org