ELPA21

English Language Proficiency Assessment for the 21st Century

Arkansas ELPA21 Scoring Interpretation Guide

Summative Assessment

Grades K-12

This guidance document is advisory in nature but is binding on an agency until amended by such agency. A guidance document does not include internal procedural documents that only affect the internal operations of the agency and does not impose additional requirements or penalties on regulated parties or include confidential information or rules and regulations made in accordance with the Administrative Procedure Act. If you believe that this guidance document imposes additional requirements or penalties on regulated parties, you may request a review of the document.

Disclaimer: The contents of this document were developed under a grant from the U.S. Department of Education. However, those contents do not necessarily represent the policy of the U.S. Department of Education and you should not assume endorsement by the Federal Government

Table of Contents

List of Tables

List of Figures

Introduction

The English Language Proficiency Assessment for the 21st Century, or ELPA21, is a test of English language proficiency. Derived from an innovative set of English language proficiency (ELP) Standards[1] developed during 2012-13, ELPA21 measures English language learners' (ELLs) 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[2] (CCSS), and the Next Generation Science Standards[3] (NGSS).

ELPA21 is founded on the belief that academic content and academic language are not distinct or separate skills. As students learn language, they simultaneously interact with grade-level academic content. Increasing the expectations for the academic content that students must master in grades K through 12 requires a parallel increase in language demands. As a result ELLs are taught (with appropriate support) the same academic content in the core subject areas (English language arts, mathematics, and science) as their classmates, at the same time they are acquiring English proficiency. As ELLs learn the academic uses of the English language, they also gain opportunity to learn the knowledge and skills necessary to be on track for college and career readiness.[4]

Introduction to the ELPA21 Assessments

ELPA21 is an evidence-centered designed (ECD) summative assessment of an ELL's language proficiency. ELPA21 is administered in winter/spring each school year (testing window is approximately January through April) to students in six grade bands: Kindergarten, grade 1, grade band 2-3, grade band 4-5, grade band 6-8, and grade band 9-12.

Comprised of innovative selected-response, constructed-response and technology-enhanced items, ELPA21 is designed to measure the four language domains of listening, reading, speaking and writing as each is embedded in the academic content expectations for, English language arts, mathematics, and science described by the CCSS and NGSS. ELPA21 summative assessments provide scale scores on each of the four domains of listening, reading, speaking and writing, which are classified into five levels of performance. Overall proficiency is determined through the pattern and level of performance across the four domains. Scale scores also are provided for each domain and overall performance and comprehension.

Braille, paper and pencil, and large-print forms are available to students who need them. Arkansas allows these forms for students whose IEP or 504 plans indicate this accommodation. For all other students, paper and pencil forms should be considered on an individual basis and require state approval. Member states collaborated throughout

[1] http://www.ccsso.org/Resources/Publications/English_Language_Proficiency_%28ELP%29_Standards_.html

[2] http://www.corestandards.org/read-the-standards/

[3] http://www.nextgenscience.org/get-to-know

[4] Although recently passed ESEA legislation uses "EL", ELPA21 uses "ELL" for consistency across documentation given adoption of "English language learner" at the beginning of the grant.

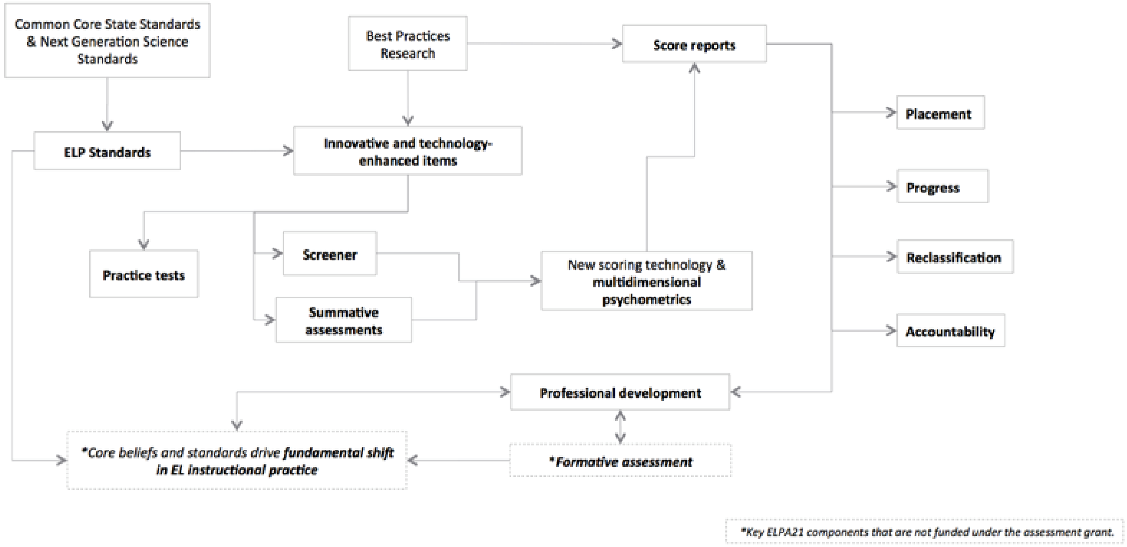
test design and development to create an accessibility framework that includes universal tools, designated supports and accommodations. These tools and supports, when used in the manner specified in the *ELPA21 Accessibility and Accommodations Manual*, ensure that ELPA21 results in valid scores for all students.

A supplemental paper form of the writing test was developed specifically to test skills and concepts that require handwriting rather than typing from our youngest students.[5] As such, all students in kindergarten and grade 1 receive two writing sections: an online portion and a paper and pencil portion.

Arkansas allows students on the ELPA21 to skip items they are unable to answer. To calibrate the item bank and to plan for a transition to computer-adaptive testing (CAT) in the future, assessments administered in 2015-16 and 2016-2017 were fixed forms that were assigned randomly to students testing within each grade band.[6]

Figure 1 describes the entire ELPA21 system; this document describes the summative assessment components only. Details of the ELPA21 screener scoring and reporting will be provided in a separate document, while the formative components not funded under the EAG Development Grant are to be developed separately.

Figure 1. ELPA21 Assessment System Diagram


Common Core State Standards & Next Generation Science Standards
Best Practices Research
Score reports
Placement
ELP Standards
Innovative and technology-enhanced items
Progress
Practice tests
Screener
New scoring technology & multidimensional psychometrics
Reclassification
Summative assessments
Accountability
Professional development
*Core beliefs and standards drive fundamental shift in EL instructional practice
*Formative assessment
*Key ELPA21 components that are not funded under the assessment grant.

*Note. *Future ELPA21 components not funded under the assessment grant.*

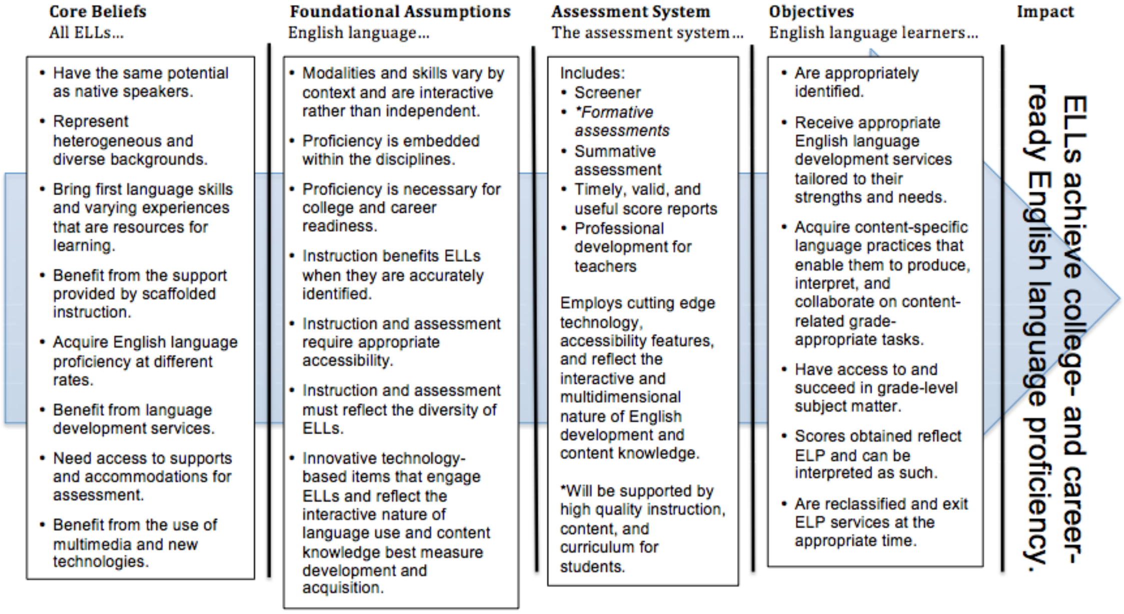
Purpose and Intended Uses of Scores

ELPA21 summative assessment results provide valuable information that informs instruction and accountability, and facilitates academic English proficiency so that all

[5] ELPA21 Assessment Guides, Grades K and 1.

ELLs have the same opportunities as their non-ELL peers to leave high school prepared for college and career success.
Figure 2 describes the theory of action that guided ELPA21 development.

Figure 2. ELPA21 Theory of Action



Core Beliefs
All ELLs...

- Have the same potential as native speakers.
- Represent heterogeneous and diverse backgrounds.
- Bring first language skills and varying experiences that are resources for learning.
- Benefit from the support provided by scaffolded instruction.
- Acquire English language proficiency at different rates.
- Benefit from language development services.
- Need access to supports and accommodations for assessment.
- Benefit from the use of multimedia and new technologies.

Foundational Assumptions
English language...

- Modalities and skills vary by context and are interactive rather than independent.
- Proficiency is embedded within the disciplines.
- Proficiency is necessary for college and career readiness.
- Instruction benefits ELLs when they are accurately identified.
- Instruction and assessment require appropriate accessibility.
- Instruction and assessment must reflect the diversity of ELLs.
- Innovative technology-based items that engage ELLs and reflect the interactive nature of language use and content knowledge best measure development and acquisition.

Assessment System
The assessment system...

Includes:
- Screener
- **Formative assessments*
- Summative assessment
- Timely, valid, and useful score reports
- Professional development for teachers

Employs cutting edge technology, accessibility features, and reflect the interactive and multidimensional nature of English development and content knowledge.

*Will be supported by high quality instruction, content, and curriculum for students.

Objectives
English language learners...

- Are appropriately identified.
- Receive appropriate English language development services tailored to their strengths and needs.
- Acquire content-specific language practices that enable them to produce, interpret, and collaborate on content-related grade-appropriate tasks.
- Have access to and succeed in grade-level subject matter.
- Scores obtained reflect ELP and can be interpreted as such.
- Are reclassified and exit ELP services at the appropriate time.

Impact

ELLs achieve college- and career-ready English language proficiency.

Note. Figure taken from ELPA21 Theory of Action, Figure 2, page 3.

The ELPA21 assessment scores serve multiple uses. They inform ELL program eligibility decisions, provide a means to monitor English proficiency progress, determine proficiency for program exit decisions, inform teachers of instructional needs of ELs, identify resource needs, and provide evidence of program effectiveness and accountability. Specifically, they are intended to meet three objectives:

- **Measuring Progress**. ELPA21 scores can be used to monitor progress by ELLs 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 ELLs and the support needs of ELL teachers, evaluating program effectiveness for subgroups of students, and adjusting educational programming and resources as needed.

- **Reclassification**. ELPA21 scores can be used to determine proficiency relative to grade appropriate performance standards for reclassification purposes. Once proficient, ELLs 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.

- **Accountability**. ELPA21 scores may be used for accountability purposes, by identifying which institutions are meeting accountability targets and which may be in need of assistance.[6]

Member states have directed the ELPA21 Consortium to provide a specific set of scores to help meet the objectives stated above; the use of these scores toward those objectives is left to each member state.

The ELP 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 ELLs learn and practice English in the classroom, they simultaneously interact with grade-level academic content. The ELP Standards describe higher expectations for ELLs by integrating language development with appropriate mathematics, English language arts, and science practices by grade. The Standards describe how language is used to meet the rigorous content demands in each grade and how ELLs progress toward English language proficiency as evidenced by:

1. Increases in the amount or sophistication of words or ways of combining words
2. Increases in repertoire of use and expansion of the types of relationships students can construct between ideas – e.g., additive, causal, conditional, contrastive – as well as the number of ways students are able to construct those relationships between ideas
3. Increases in accuracy in constructing precise meanings
4. Increases in contextualization, the ability to tailor the use of language functions to fit a variety of sociocultural contexts
5. Increases in autonomy, which is observed by the need for fewer language supports and scaffolds as proficiency increases

The ten ELP Standards can be grouped by modality and domain (Table 1). The domains are also referred to as reading comprehension, written production, listening comprehension, and oral production skills.

[6] As determined by the US Department of Education's current accountability legislation.

Table 1. ELP Standards, Modalities, and Domains

Feature	Standard #	ELP Standard	Modality	Domain			
				L	**R**	**S**	**W**
Language Necessary for Engagement in Content Area Practices	3	speak and write about grade-appropriate complex literary and informational texts and topics	Productive			✓	✓
	4	construct grade-appropriate oral and written claims and support them with reasoning and evidence					
	7	adapt language choices to purpose, task, and audience when speaking and writing					
	2	participate in grade-appropriate oral and written exchanges of information, ideas, and analyses, responding to peer, audience, or reader comments and questions	Interactive	✓	✓	✓	✓
	5	conduct research and evaluate and communicate findings to answer questions or solve problems					
	6	analyze and critique the arguments of others orally and in writing					
	1	construct meaning from oral presentations and literary and informational text through grade-appropriate listening, reading, and viewing	Receptive	✓	✓		
Micro-linguistic Features	8	determine the meaning of words and phrases in oral presentations and literary and informational text					
	9	create clear and coherent grade-appropriate speech and text	Standards 9 and 10 address the linguistic structures of English and are framed in relation to the CCSS for ELA Language domain.				
	10	make accurate use of standard English to communicate in grade appropriate speech and writing					

Note. Because the ability to communicate via multiple modes of representation (e.g., non-verbal communication, oral, pictorial, graphic, textual) may be especially important for ELLs with certain types of disabilities, ELPA21 carefully considered the access supports and accommodations for ELLs with IEPs or 504 plans.

The ELP Standards further differentiate basic conversational interaction (Standard 2) from interaction with academic content knowledge that requires higher order thinking (Standards 5 & 7). Production could be divided by domain (speaking and writing), but also by standard (Adaptive Production (standard 7) and Basic Production (Standards 3 & 4.)

The 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.

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.

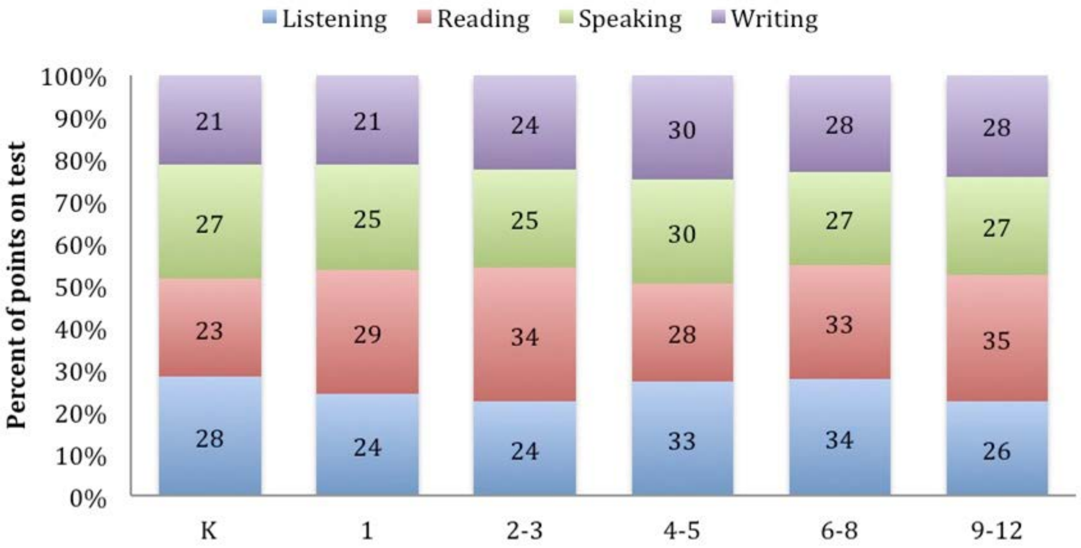
The receptive modality refers to the learner as a reader and listener/viewer working with ‘text’ whose author or deliverer is not present or accessible. It presumes authentic written or oral documents where language input is meaningful and content laden. The receptive modality requires skills necessary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

The interactive modality emphasizes the need for ELLs 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.

The Domains

The four language domains of listening, reading, speaking and writing are contained within the three modalities. ELPA21 measures and reports English language proficiency using these domains. On the summative assessments, the number of score points and items are relatively evenly distributed across domain and grade band (Figure 3.)

Figure 3. Domain Representation as Percent of Total Points by Grade Band

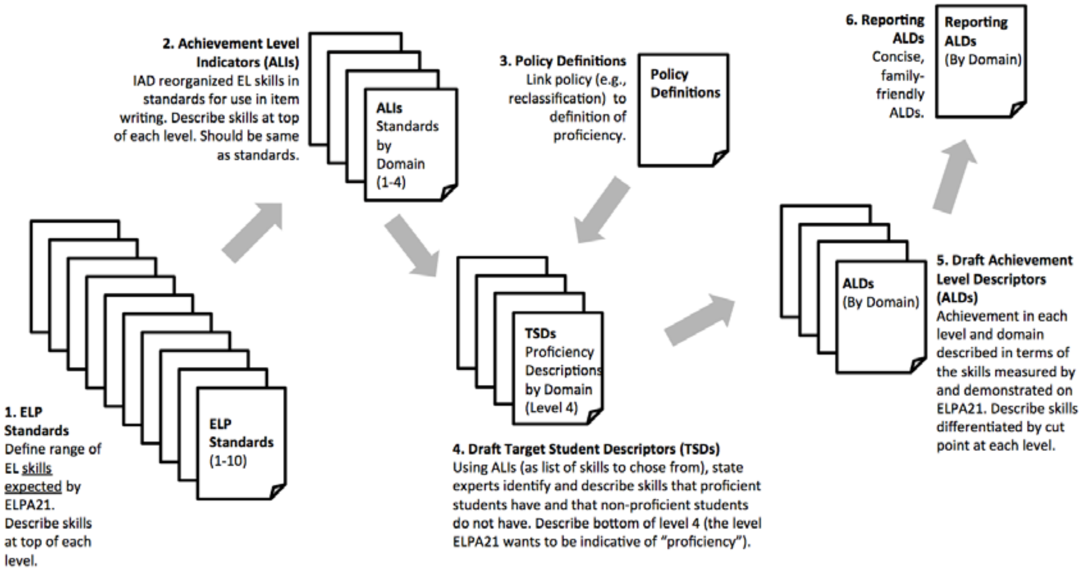


Source: ELPA21 Assessment Framework, Table 5.1, ELPA21 Operational Summative Assessment Test Blueprints.
Notes. Tests range in number of score points from 99-123. Number of points per grade band is 99 for K and grade 1, 107 for grades 2-3, 121 for grades 4-5, 123 for grades 6-8, and 116 for grades 9-12.

Describing Proficiency

ELPA21 has multiple definitions of proficiency, each serving a specific purpose.

Figure 4. Descriptions of Proficiency


2. Achievement Level Indicators (ALIs)
IAD reorganized EL skills in standards for use in item writing. Describe skills at top of each level. Should be same as standards.
ALIs Standards by Domain (1-4)
3. Policy Definitions
Link policy (e.g., reclassification) to definition of proficiency.
Policy Definitions
6. Reporting ALDs
Concise, family-friendly ALDs.
Reporting ALDs (By Domain)
1. ELP Standards
Define range of EL skills expected by ELPA21. Describe skills at top of each level.
ELP Standards (1-10)
TSDs Proficiency Descriptions by Domain (Level 4)
4. Draft Target Student Descriptors (TSDs)
Using ALIs (as list of skills to chose from), state experts identify and describe skills that proficient students have and that non-proficient students do not have. Describe bottom of level 4 (the level ELPA21 wants to be indicative of "proficiency").
ALDs (By Domain)
5. Draft Achievement Level Descriptors (ALDs)
Achievement in each level and domain described in terms of the skills measured by and demonstrated on ELPA21. Describe skills differentiated by cut point at each level.

For clarity in wording through this document, Table 2 provides ELPA21's chosen terminology for the different definitions of proficiency referenced throughout this document.

Table 2. Types of Performance Descriptors

ELPA21 Terminology	General Use	General Audience	General Type
ELP Standard PLDs	Described in ELP Standards. Generally include five levels of performance, although some knowledge, skills, and abilities are specific to low or high performance on individual standard	Educators, item writers, curriculum developers	PLDs by standard
Policy Definitions (PDs)	Describe the rigor of, and ELPA21's vision for, English language proficiency and its impact on policy, consistent across grade.	Policy-makers	Policy-Level PLDs
Achievement Level Indicators (ALIs)	Also known as *range PLDs, item writing PLDs*, or as *developmental* or *learning trajectories* of English language proficiency, they reorganize the ELP Standards by domain.	Educators, item writers	Range PLDs
Target Student Descriptions	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.	Standard setting panels	Interim ALDs

Achievement Level Descriptors (ALDs)	Target PLDs reflect cut scores established through standard setting. They characterize the knowledge and skills differentiating the performance levels from each other by describing the minimal knowledge, skills, and abilities required for each level.	Stakeholders, score report audiences	Final ALDs
Reporting PLDs	Derived from the final PLDs, they describe the appropriate inferences that may be made about the students who score in each performance level.	Score report audiences	Reporting PLDs

Policy Definitions

Performance levels are broad categories that describe the results of an assessment, and ELPA21 has five. These levels describe the stages of English language development through which ELLs are expected to progress as they gain proficiency (Table 3.)

Table 3. Policy Definitions for the Five ELPA21 Performance Levels

A STUDENT AT THIS LEVEL…	
Level 1: Beginning	Displays **few** grade-level English language skills and will benefit from EL Program support.
Level 2: Early Intermediate	Presents evidence of **developing** grade-level English language skills and will benefit from EL Program support.
Level 3: Intermediate	Applies **some** grade-level English language skills and will benefit from EL Program support.
Level 4: Early Advanced	Demonstrates English language skills **required for engagement** with grade-level academic content instruction at a level comparable to non-ELs.
Level 5: Advanced	Exhibits **superior** English language skills, as measured by ELPA21.

Note[7]: Definitions assume proficiency is demonstrated by scoring just above the cut score between Levels 3 and 4, or are in the bottom of the range of skills described by Level 4.

Performance Targets

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". Table 4 describes the performance target for each of the four domains.

Table 4. Performance Targets by Domain

DOMAIN	**DEFINITION**
ELs demonstrate skills required for engagement with grade-level academic content instruction at a level comparable to non-ELs. For each domain…	
Listening	An EL can listen and comprehend **spoken English** at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.

[7] The final number of levels will be verified via the operational distribution of students across the five levels; some levels may ultimately be combined.

Reading	An EL can read and comprehend **written English** at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Speaking	An EL can **produce speech** at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.
Writing	An EL learner can **write texts** at a level sufficient to fully participate in and learn from grade-level instruction, communication, and activities.

Achievement Level Descriptors

ALDs describe student's actual performance on the summative assessment for each of five levels. They are finalized following standard setting and are available on NDE's Title III website under ELPA21 Assessment:
https://www.education.ne.gov/NATLORIGIN/ELPA21.html

The Scoring Model

In its role as lead psychometrics, validity, and scoring vendor for ELPA21, the *National Center for Research on Evaluation, Standards, and Student Testing* (CRESST) determines the scoring model, the associated psychometrics and the evidence necessary for establishing validity.

Test Scoring

Documentation of the approach is available in the *ELP21 2016 Summative Assessment Scoring and Scaling Specifications*.

Item Scoring

ELPA21 is built around a set of machine- and hand-scored task types (Table 5.) Some tasks are specific to individual grade levels or domains, while others apply to all grade bands and domains.

Table 5. Task Types by Domain

Listening	**Reading**	**Speaking**	**Writing**
• Academic Debate • Academic Lecture and Discussion • Academic Lecture or Discussion • Follow Instructions • Interactive Student Presentation • Listen and Match • Listen for Information • Long Conversation • Read-Aloud Story • Short Conversations • Student Discussion • Teacher Presentation	• Argument and Support Essay Set • Discrete Items • Informational Set • Short Informational Set • Extended Informational Set • Literary Set • Short Literary Set Extended Literary Set • Short Literature Set • Extended Literature Set • Match Picture to Word and Sentence • Procedural Text • Read and Match	• Academic Debate • Analyze a Visual and a Claim • Classroom Tableau • Conversation • Language Arts Presentation • Observe and Report • Opinion • Picture Description/Compare Pictures • Read Aloud • Show and Share • Student Discussion	• Complete a Word • Construct a Claim • Copy a Word • Opinion • Picture Caption • Respond to Peer Email • Storyboard • Write a Sentence • Write a Word • Writing Questions

• Teacher Presentation: Read Aloud	• Read for Details • Read-along Sentence • Read-Along Story • Short Correspondence • Short Correspondence Set • Word Wall		

Source: ELPA21 Item Development Process Report FINAL ETS Submission 5-15-2015-1.pdf and ELPA21 Assessment Frameworks, Documents are available in ELPA21 Operational Handoffs folder.

Tasks contain technology-enhanced items (TEI), selected response (SR) items, and constructed response (CR) items. Reading and listening tests contain machine-scored selected response (SR) and technology-enhanced items (TEI). Writing tests contain both hand- and machine-scored SR and CR items, and speaking tests contain all constructed response (CR) items. The proportion of machine-scored to hand-scored items decreases as the grade level increases.

ELPA21 requires that member states utilize centralized scoring, and allows for local scoring of constructed responses only when state policy allows, such as when a student's documented cultural practices prohibit the use of technology to capture their responses. Documentation of the scoring process is available in *ELPA21 Partial Credit Scoring Rules Validation Report* and *ELPA21 Hand-scoring Rubrics* and item-specific rubrics are contained in the item XML. Rules related to hand-scoring are listed in Table 6.

Table 6. Hand-Scoring Rules

Rule
• Some drag-and-drop items contain more objects (to be dragged) and "drop zones" (where objects are dragged to) than were needed to respond to the item correctly. No penalty shall be applied to students who provide extraneous responses by dragging objects to additional locations.[8]
• On some items, students might receive partial credit by simply following the directions for an item. If it is possible to get one (or two) response(s) correct simply by completing the task, the scoring rule is to not provide automatic credit.[9]
• Sequence items require students to put information into a correct sequence and will receive credit for fully correct sequences only. For example, if a student places objects into order 1,3,2,4, she would be given partial credit for correctly sequencing the first and final part of the item. However, if a student placed the sources into order 2,3,4,1, zero points (no partial credit) would be awarded based on the preliminary scoring rules, even though there is a partial order string (2,3,4) in the correct sequence. [10]

Note: All rules are in item scoring XML and require no manual application.

Scores

ELPA21 provides scores to be used for reporting and include a summary of performance on the four domains and a Proficiency Determination of Emerging, Progressing, and Proficient that is based on the pattern (or profile) of performance across the four domains.

[8] ELPA21 Assessment Framework – Summative, page 54
[9] ELPA21 Assessment Framework – Summative, page 56
[10] ibid

Arkansas reports domain performance as a scale score and a level. These scores are provided for use by students, educators, and parents and meet the ELPA21 objectives of measuring progress and determining program eligibility.

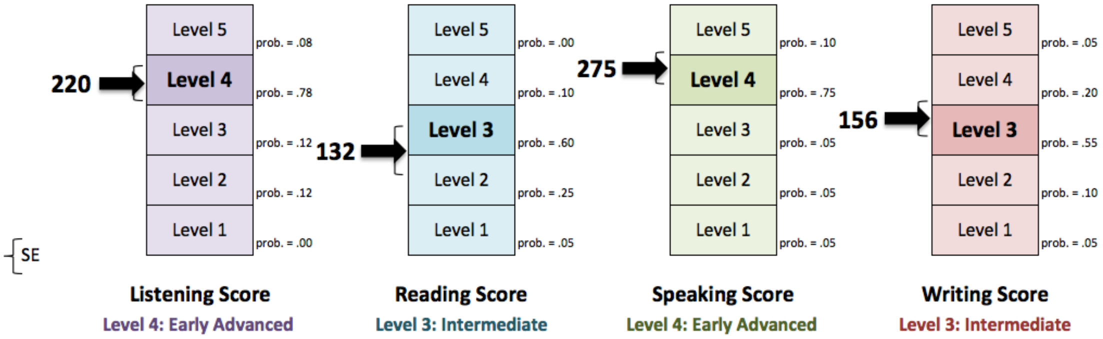
Reporting Scores

The Domain Scores

A numeric three-digit scale score describes performance on the four domains of listening, reading, speaking and writing. 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 *ELP21 2016 Summative Assessment Scoring and Scaling Specifications* and the corresponding descriptors (ALDs), available at https://www.education.ne.gov/NATLORIGIN/ELPA21.html

A measure of variability describes the variability (standard error) around each numeric score, and a probability of classification describes the likelihood of classification into each of the five levels. The closer a domain score is to a cut score, the larger the probability will be that the student's true score falls into the adjacent category. For example, the true score for a student scoring 132 has a higher probability of falling into Level 3 or Level 2 if the cut between level 2 and level 3 is set at 131 than if it were set at 119 (see Figure 5).

Figure 5. The Domain Scores



Note. Level 4 represents the assessment target for each domain.

Although performance on a single domain cannot determine English language proficiency, the question "what level on each domain is sufficient?" begs to be addressed. **Corresponding to the policy definition, 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-ELLs 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.**

The Profiles

The ELP Standards maintain that proficiency can be achieved in multiple ways, and may look different for individual students. Students develop skill in each domain at different rates and may exhibit some skills of a domain at higher levels and still struggle with other skills at a lower level. As such, ELPA21 recognizes the possibility of other profiles of skills across the domains of listening, reading, speaking and writing that may describe proficiency in addition to the “4444” profile.

Profiles may be expressed as 1) four numbers representing the level of proficiency on each of the four domains, or as 2) rules that summarize a common pattern for sets of profiles. As an example, the hypothetical scores described in Figure 5 could be expressed as a profile of “4343” or by the rule “no domain score falls below Level 3”. Profiles are based on the domain scores of proficient students identified during the ELPA21 contrasting group study, were refined and vetted by a panel of member state EL experts, and were refined and vetted again during standard setting.

Relying on student profiles instead of an overall composite score as is traditionally done benefits educators in a couple ways. A profile provides more instructional information about students who may have the same overall score, but differ in skills and needs. Proficiency is a function of the domains, not of the overall scores with domains combined. It also highlights the relationship between the domains in a way that an overall score would not.

Profiles are used to differentiate “Proficient” students from those who are “Progressing” or “Emerging”. Table 7 describes how the different profiles are expressed and used to determine proficiency on ELPA21.

Table 7. Profiles of Proficiency

Rules	**Profiles (examples)**	**Proficiency Determination**
A profile of 4s and 5s meets assessment targets and indicates overall proficiency	4444 5555 4545 5454 4455 5544 4445 4454 4544 5444 5554 5545 5455 4555 4E44	Proficient
A profile with one or more domain scores above Level 2 that does not meet the requirements to be Proficient	3333 1333 3353 3233 2242 1234 1114 2232	Progressing
A profile of 1s and 2s indicates an “Emerging” level of proficiency.	1122 1212 E222 2222	Emerging

Note. The order of the example profiles of the four domains is: 1) reading, 2) writing, 3) speaking and 4) listening. “E” indicates an exempt test.

The Proficiency Determination

Using the profiles, different combinations of skills and abilities across the domains are deemed as “Proficient”, “Progressing” or “Emerging” (see Table 8). The Proficiency Determination (often referred to by states as the Overall Proficiency Determination) identifies ELLs whose language skills enable full participation in grade-level academic contexts.

Table 8. Policy Definition for the Proficiency Determination

Proficient	Students are Proficient when they attain a level of English language skill necessary to independently produce, interpret, collaborate on, and succeed in grade-level content-related academic tasks in English. This is indicated on ELPA21 by attaining a profile of Level 4 or higher in all domains. Once Proficient on ELPA21, students can be considered for reclassification.
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.
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.

Note. Each definition consists of three elements: a learning expectation (first sentence), an operational definition (second sentence), and a policy impact statement (third sentence).

Using profiles of proficiency allows for limited compensation within a mostly conjunctive classification model. Defining proficiency as profiles of skills rather than as an average or a sum across those skills allows for richer, more nuanced and flexible distinctions between proficient and not proficient students. For example, a profile of “4131” is more useful and provides educators with more information than does an overall composite score of 360. The profile, while not diagnostic, does show clear strength and weakness.

Note that summative assessments are developed by grade-band, but scores are reported at grade-level, and as a result, different profiles may indicate proficiency in different grades. Because of this, students at different grades within the same grade-band who earn identical scores may fall into different achievement levels and receive different proficiency determinations. This is because the expectation (e.g., cut score, or standard for proficiency) increases for each grade. For example, referring back to Figure 5, a student receiving a 220 in listening, a 132 in reading, a 275 in speaking and a 156 in writing may be Progressing (a profile of 4343) in 8th grade, but Proficient in 6th grade.

Business Rules

The ELPA21 business rules define or constrain some aspect of collecting, scoring, manipulating, or reporting scores from ELPA21 summative assessments. These eight rules are intended to meet two objectives of the grant: 1) implement a shared definition of proficient across member states, and 2) ensure score comparability across member states.

Attempted Test Rules

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).

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.

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.

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 (see the Missing Responses rule above).

Exempted Test Rules

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"). An Arkansas district must request an exemption approval from the state.

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 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.

Score Reporting

Every ELPA21 state designs, creates, and distributes score reports to a variety of audiences (Table 9.)

Table 9. Score Reports and Audiences

Audience	Report
Students & Family	Individual Student Report
Teachers	Teacher/Classroom Summary Report Individual Student Reports
Teachers and School Administrators	Teacher/Classroom Summary Reports School Aggregate Report
District Administrators	District Aggregate Report
State Administrators and Policy Makers	State Aggregate Report

Individual Student Report - Arkansas

The Individual Student Report (ISR), provided to students and their parents or guardians, includes:

1. A determination of overall proficiency
2. Performance level and descriptor for each domain score
3. Scale scores for each of the four domains
4. Explanatory text for concepts that may be easily understood or not commonly known

Roster Score Reports

Arkansas provides roster reports to school administrators and educators summarizing their students' performance. Member states indicated ELLs 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, the district and school roster reports are arranged by grade level.

Aggregate Score Reports

States will provide aggregate score reports to school administrators and educators summarizing their students' performance grade-level and other groups as determined by state policy. Aggregate reports (called demographic summary reports) for Arkansas contain the following subgroups:

1) Gender
2) Ethnicity
3) IEP status
4) ELL Status
5) Economic situation

General Score Interpretation Guidance

ELPA21 offers unique information to students, parents, educators, administrators, and policy-makers. When used appropriately, this information describes what ELLs know and can do in terms of the grade-level language skills required to engage with the content that is taught according to rigorous academic standards.

ELPA21, like all tests, has limitations. No single test can measure all aspects of a student’s language use, and no test can measure this perfectly. ELPA21 scores are provided with a measure of error. Summative assessment scores should be viewed as one indicator among multiple sources of evidence (such as classroom-based tests, course grades, teacher observations, and samples of student work) when interpreting and making decisions about a student’s English language proficiency.

When used as designed, ELPA21 provides useful information. However, like any other test, it may have unintended consequences if used outside the specific purposes and populations for which it was designed and validated.

ELPA21 uses valid psychometric processes to ensure that scores from different test forms describe the same level of performance. For example, score from a 5th grade student scoring just above proficient on test form A and a 5th student scoring just above proficient on form B would represent the same performance, and these scores are comparable. When aggregated, these scores can also describe school- or district-level changes in proficiency and can measure gaps in achievement among different groups of students.

ELPA21 tests scores should not be compared to scores from any state’s previous ELP exam. The implementation of the ELP standards generally results in the development of new curriculum and instructional strategies, making results across the tests incomparable.

Decisions based on these scores should follow a process established by each district that includes other sources of information such as teacher feedback, grades, etc.

Scores within a single grade band are comparable. Those across grade bands are not.

Key Definitions

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

Calibration: To set or establish 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 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 response.

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.

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.

Constructed-Response Item/Constructed Response (CR): A type of item on ELPA21 requiring a student response that is in a written, typed, spoken, or action format (e.g., short answer, essay, research report, oral presentation, demonstration). The terms open-ended and free-response are often used interchangeably with constructed-response.[11]

Content-specific[12]: 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.

[11] Council of Chief State School Officers (2013). Operational Best Practices for Statewide Large-Scale Assessment Programs.

[12] The ELP Standards, page 211

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.[13]

English Language Learner (ELL): According to the USDOE, an English language 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
III. The opportunity to participate fully in society.[14]

ELPA21 applies this definition to defining ELLs.

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.[15]

Evidence-Centered Design (ECD): An approach, followed by ELPA21, to constructing educational assessments that utilizes evidentiary reasons and arguments[16]. 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)[17]

[13] http://ncme.org/resource-center/glossary/
[14] Public Law 107-110. Title IX, Part A, Sec. 9101, (25)
[15] See Harvill, L. M. (1991), Standard Error of Measurement. Educational Measurement: Issues and Practice, 10: 33–41
[16] 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
[17] Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23

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.)[18]

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.

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. ELPA21 tests are high stakes.

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

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.[19]

Paper-based/Paper-and-pencil (p&p): A form of ELPA21 delivered in a printed hard copy form, rather than in a digital form.[20]

Performance-Level Descriptors (PLDs): A general term referring to multiple descriptions of what performance at each level of the test should and does look like. In score reporting, they describe what scores mean and communicate what students scoring at each level know and are able to do.

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

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.[21]

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

[18] The ELP Standards, page 213
[19] Ibid
[20] Council of Chief State School Officers (2013). Operational Best Practices for Statewide Large-Scale Assessment Programs.
[21] Assessing Students with Disabilities: A Glossary of Assessment Terms in Everyday Language

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).[22]

Scale Score: A kind of score to which a raw score has been converted to a numeric scale for ease of interpretation.[23]

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 ELL services.

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.

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”.[24]

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).[25]

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.[26]

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.[27]

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

[22] ibid
[23] http://ncme.org/resource-center/glossary/
[24] http://ncme.org/resource-center/glossary/
[25] Ibid
[26] Ibid
[27] Assessing Students with Disabilities: A Glossary of Assessment Terms in Everyday Language

secure test materials, and/or the established standardized test administration protocols.[28] Test irregularities may invalidate test scores that were obtained during the irregularity.

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

[28] Council of Chief State School Officers (2013). Operational Best Practices for Statewide Large-Scale Assessment Programs.

References

Abedi, J. (2002). *Standardized achievement tests and English language learners: Psychometric issues.* Educational Assessment, 8, 231-257.

Abedi, J. (2007). *English language proficiency assessment in the nation: Current status and future practice.* Davis: University of California.

Abedi, J. (2008a). Measuring Students' Level of English Proficiency: Educational Significance and Assessment Requirements. *Educational Assessment*, (13)193-214.

Abedi, J. (2008b). Classification System for English Language Learners: Issues and Recommendations. *Educational Measurement, Issues and Practice.* 17-31.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, D.C.: Authors.

American Institutes for Research. (2015). *Smarter Balanced Scoring Specification, 2014-2015 Administration. Summative and Interim Assessments: ELA Grades 3-8, 11 and Mathematics, Grades 3-8, 11.* Version 7.

Briggs, D. C. & Weeks, J. P. (2009). The impact of vertical scaling decisions on growth interpretations. *Educational Measurement: Issues and Practice, 28(4), 3-14.*

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests.* Thousand Oaks, CA: Sage.

Council of Chief State School Officers. (2014). *English Language Proficiency (ELP) Standards with Correspondences to K–12 English Language Arts (ELA), Mathematics, and Science Practices, K–12 ELA Standards, and 6-12 Literacy Standards.* Washington, DC: CCSSO.

Council of Chief State School Officers. (2013). *Operational Best Practices for Statewide Large-Scale Assessment Programs.* Washington, DC: CCSSO.

Custer, M., Omar, M. H. & Pomplun, M. (2006). Vertical scaling with the Rasch model utilizing default and tight convergence settings with WINSTEPS and BILOGMG. *Applied Measurement in Education, 19(2), 133-149.*

Duran, R. P. (2008). Assessing English-Language Learners' Achievement. *Review of Education*, (32)292-327.

Egan, K. L., Ferrara, S., Schneider, C. M., Barton, K. E. (2009). Writing Performance Level Descriptors and Setting Performance Standards for Assessments of Modified

Achievement Standards: The Role of Innovation and the Importance of Following Conventional Practice. *Peabody Journal of Education*, 84(4).

English Language Proficiency Assessment for the 21st Century (2014). *Theory of Action*.

Faulker-Bond, M., Wolf, M. K., Wells, C. S., and Sireci, S. G. (no date). *Exploring the Factor Structure of a K-12 English Language Proficiency Assessment*.

Hakuta, K., & Beatty, A. (Eds.). (2000). *Testing English language learners in U. S. schools: Report and workshop summary*. Washington, DC: National Academy Press.

Hakuta, K. (2011). Educating language minority students and affirming their equal rights: Research and practical perspectives. *Educational Researcher, 40*(4), 163-174.

Hambleton, R. K., & Pitoniak, M. (2006). *Setting performance standards*. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 433-470). Westport, CT: Praeger.

Hattie, J., (1985). Methodology Review: Assessing Unidimensionalty of Tests and Items. *Applied Psychological Measurement*, (9),(2) 139-164.

Hauck, M. C., Pooler, E., and Anderson, D. P. (2015). *ELPA21 item development process report*. Report submitted by Educational Testing Service (ETS), May 15, 2015.

Kolen, M. J. & Brennan, R. L. (2004). *Test equating, scaling, and linking: methods and practices (2nd edition).* New York: Springer.

Linquanti, R., & Cook, G. (2013a). *Toward a "common definition of English learner": A brief defining policy and technical issues and opportunities for state assessment consortia.* Washington DC: CCSSO. Retrieved July 23, 2013, from http://www.ccsso.org/Documents/2013/Common%20Definition%20of%20English%20Learner_2013.pdf.

Linquanti, R., & Cook, G. (2013b). *Toward a "common definition of English learner": Guidance for States and State Assessment Consortia in Defining and Addressing Policy and Technical Issues and Options.* Washington DC: CCSSO.

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

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

Perie, M. (2006). *Convening an articulation panel after a standard setting meeting: A how-to guide.*

Perie, M. (2008). A guide to understanding and developing performance-level descriptors. *Educational Measurement: Issues and Practice, 27(4), 15-29.*

Rabinowitz S., & Sato E. (2006, April). Technical adequacy of assessments for alternate student populations: Technical review of high-stakes assessment for English language learners. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Reckase, M. D., (2008). *Study of Best Practices for Vertical Scaling and Standard Setting, with Recommendations for FCAT 2.0.*

Schneider, M. C., & Egan, K. L. (no date). *A Handbook for Creating Range and Target Performance Level Descriptors.* The National Center for the Improvement of Educational Assessment.

Tippeconnic, J. W., III, & Faircloth, S. C. (2002). Using culturally and linguistically appropriate assessments to ensure that American Indian and Alaska Native students receive the special education programs and services they need (EDO-RC-02-8). Charleston, WV: ERIC Clearinghouse on Rural Education and Small Schools.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service, National Evaluation of Title III Implementation Supplemental Report: Exploring Approaches To Setting English Language Proficiency Performance Criteria and Monitoring English Learner Progress , Washington, DC 2012.

Wolf, M. K., Farnsworth, T. & Herman, J. (2008). Validity Issues in Assessing English Language Learners' Language Proficiency. *Educational Assessment*, 13:80-107.

Wolf, M. K., Kao, J. C., Herman, J. L., Bachman, L. F., Bailey, A. L., Bachman, P. L., et al. (2008a). Issues in assessing English language learners: English language proficiency measures and accommodation uses—Literature review (CRESST Tech. Rep. No. 731). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Wolf, M. K., Kao, J. C., Griffin, N., Herman, J. L., Bachman, P. L., Chang, S. M., et al. (2008b). Issues in assessing English language learners: English language proficiency measures and accommodation uses—Practice review (CRESST Tech. Rep. No. 732). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Wolf, M. K., Kao, J. C., Herman, J. L., Bachman, L. F., Bailey, A. L., Bachman, P. L., et al. (2008c). Recommendations for Assessing English Language Learners: English Language Proficiency Measures and Accommodation Uses (CRESST Tech. Rep. No. 737). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Zieky, M. J., Perie, M., & Livingston, S. (2008). *Cutscores: A manual for setting standards of performance on educational and occupational tests.* Princeton, NJ: Educational Testing Service (ETS).