

2.4 ACCESS Online Score Scale Maintenance: Writing

Series 400 to Series 303

Prepared by:
Dorry Kenyon and Shu Jing Yen
October 2015

Issue

The purpose of this document is to provide an overview of the planned approach to linking the Grades 1-12 ACCESS for ELLs 2.0 Series 400 Online (hereafter ACCESS 2.0) assessment of academic English language proficiency development expressed in the Writing domain (hereafter “Writing Test”) to the ACCESS for ELLs (hereafter ACCESS) Writing Test in order to maintain the ACCESS Writing reporting scale.

Background

The ACCESS 2.0 Writing domain is designed to measure the same construct as the ACCESS Writing domain. To this end, the design of ACCESS 2.0 Writing shares these features with ACCESS Writing:

- All Writing tasks are constructed response.
- Prompts are written to address one or more of the five WIDA English language development (ELD) Standards. Most test forms will have three prompts, addressing the following standards:
 - Prompt 1: Social and Instructional Language
 - Prompt 2: Integrated Language of Math/Language of Science
 - Prompt 3: Integrated Language of Language Arts/Language of Social Studies/Social and Instructional Language
- Tasks are designed to allow students to show what they can do with their developing academic English language proficiency in writing. The tasks provide rich content so that prior content knowledge is not needed to write a response.

Although the overall test design remains unchanged in ACCESS 2.0, there are several minor changes in how Writing tasks are administered, answered, and scored. Additionally, the tier structure and the elementary grade-level clustering has changed for ACCESS 2.0. These changes are intended to have a minimal impact on how the Writing domain is measured; however, they carry some implications for how the ACCESS Writing Test scale can be maintained. Table 1 compares the ACCESS 2.0 Writing Test and the ACCESS Writing Test. Differences in the prompt delivery and response mode between ACCESS 2.0 and ACCESS are summarized below. Differences in the scoring rubric between ACCESS 2.0 and ACCESS are explained in detail in the subsection *ACCESS Writing Rating Scale Maintenance*.

Table 1

Comparison between the ACCESS 2.0 Writing Test (Series 400) and the ACCESS Writing Test

	ACCESS 2.0 (Series 400)	ACCESS
Grade Clusters	1 2-3 4-5 6-8 9-12	1-2 3-5 6-8 9-12
Tier Structure	A, B/C	A, B, C
Item Specification	Same as ACCESS	
Item Content	Adapted from Old ACCESS Tasks	
Prompt Delivery	Grades 1-3: Paper Booklet Grades 4-12: Computer	Paper Booklet
Response Mode (Default)	Grades 1-3 : Handwritten Grades 4-5: Keyboarded/Handwritten* Grades 6-12 Keyboarded	Handwritten
Tasks	Same as ACCESS	Three Tasks (Two Short Tasks and One Long Task)
Scoring Rubrics	Revision in Rubrics and Rating to Raw Score Conversion	

* State decision regarding default option

Delivery of Prompts

While the Writing prompts on ACCESS are delivered in a printed booklet, the prompts for ACCESS 2.0 Writing are delivered on the computer screen for Grades 4-12; the Grade 1 and Grades 2-3 clusters will continue to be delivered on paper. Every effort was made through cognitive labs and small-scale tryouts to ensure that displaying the Writing prompts on the computer screen would not cause difficulties for students.

Responses

On ACCESS, all Writing responses are handwritten. On ACCESS 2.0, responses can be handwritten or keyboarded. In grades 4-5, the default Writing response mode is either handwritten or keyboarded, which is at the discretion of the state. For all students in Grades 6-8 and Grades 9-12, keyboarding is the default mode. Nevertheless, in all grades, the option to write by hand may be provided to individual students as an accommodation.

ACCESS Writing Rating Scale Maintenance

With the development of ACCESS 2.0, the WIDA Writing Rubric was updated to reflect the amplified performance level definitions of the 2012 publication. The background on the original

WIDA Writing Rubrics as well as the revisions made to maintain the Writing rating scale are described below.

Background on the Original WIDA Writing Rubrics

The original purpose of ACCESS was to operationalize the WIDA ELD Standards, such that scores awarded to performances on the assessment could be related to the WIDA Standards in a straightforward way. To this end, the original WIDA Writing Rubric was developed to directly reflect the five performance level definitions in the original WIDA Standards (2004), shown in Figure 1.

At the given level of English language proficiency, English language learners will process, understand, produce, or use:	
5- Bridging	<ul style="list-style-type: none"> • the technical language of the content areas; • a variety of sentence lengths of varying linguistic complexity in extended oral or written discourse, including stories, essays, or reports; ➢ oral or written language approaching comparability to that of English proficient peers when presented with grade level material
4- Expanding	<ul style="list-style-type: none"> • specific and some technical language of the content areas; • a variety of sentence lengths of varying linguistic complexity in oral discourse or multiple, related paragraphs; ➢ oral or written language with minimal phonological, syntactic, or semantic errors that do not impede the overall meaning of the communication when presented with oral or written connected discourse with occasional visual and graphic support
3- Developing	<ul style="list-style-type: none"> • general and some specific language of the content areas; • expanded sentences in oral interaction or written paragraphs; ➢ oral or written language with phonological, syntactic, or semantic errors that may impede the communication but retain much of its meaning when presented with oral or written, narrative or expository descriptions with occasional visual and graphic support
2- Beginning	<ul style="list-style-type: none"> • general language related to the content areas; • phrases or short sentences; ➢ oral or written language with phonological, syntactic, or semantic errors that often impede the meaning of the communication when presented with one to multiple-step commands, directions, questions, or a series of statements with visual and graphic support
1- Entering	<ul style="list-style-type: none"> • pictorial or graphic representation of the language of the content areas; • words, phrases, or chunks of language when presented with one-step commands, directions, WH-questions, or statements with visual and graphic support

Figure 1. Performance Level Definitions from the 2004 WIDA ELD Standards

To aid raters, the original WIDA Writing Rubric organized its descriptors under three categories: linguistic complexity, vocabulary usage, and language control. The rubric also contained definitions for a Performance Level 6, “Reaching” (see Figure 2). This performance level was alluded to but not defined in the 2004 WIDA ELD Standards, and was subsequently added in the 2007 revised WIDA ELD Standards (see Figure 3).

Writing Rubric of the WIDA™ Consortium Grades 1-12			
Level	Linguistic Complexity	Vocabulary Usage	Language Control
6 Reaching*	A variety of sentence lengths of varying linguistic complexity in a single tightly organized paragraph or in well-organized extended text; tight cohesion and organization	Consistent use of just the right word in just the right place; precise Vocabulary Usage in general, specific or technical language.	Has reached comparability to that of English proficient peers functioning at the “proficient” level in state-wide assessments.
5 Bridging	A variety of sentence lengths of varying linguistic complexity in a single organized paragraph or in extended text; cohesion and organization	Usage of technical language related to the content area; evident facility with needed vocabulary.	Approaching comparability to that of English proficient peers; errors don't impede comprehensibility.
4 Expanding	A variety of sentence lengths of varying linguistic complexity; emerging cohesion used to provide detail and clarity.	Usage of specific and some technical language related to the content area; lack of needed vocabulary may be occasionally evident.	Generally comprehensible at all times, errors don't impede the overall meaning; such errors may reflect first language interference.
3 Developing	Simple and expanded sentences that show emerging complexity used to provide detail.	Usage of general and some specific language related to the content area; lack of needed vocabulary may be evident.	Generally comprehensible when writing in sentences; comprehensibility may from time to time be impeded by errors when attempting to produce more complex text.
2 Beginning	Phrases and short sentences; varying amount of text may be copied or adapted; some attempt at organization may be evidenced.	Usage of general language related to the content area; lack of vocabulary may be evident.	Generally comprehensible when text is adapted from model or source text, or when original text is limited to simple text; comprehensibility may be often impeded by errors.
1 Entering	Single words, set phrases or chunks of simple language; varying amounts of text may be copied or adapted; adapted text contains original language.	Usage of highest frequency vocabulary from school setting and content areas.	Generally comprehensible when text is copied or adapted from model or source text; comprehensibility may be significantly impeded in original text.

Figure 2. Original Writing Rubric of the WIDA Consortium

WIDA Performance Definitions

At the given level of English language proficiency, English language learners will process, understand, produce or use:

6- Reaching	<ul style="list-style-type: none"> specialized or technical language reflective of the content areas at grade level a variety of sentence lengths of varying linguistic complexity in extended oral or written discourse as required by the specified grade level oral or written communication in English comparable to English-proficient peers
5- Bridging	<ul style="list-style-type: none"> specialized or technical language of the content areas a variety of sentence lengths of varying linguistic complexity in extended oral or written discourse, including stories, essays or reports oral or written language approaching comparability to that of English-proficient peers when presented with grade level material
4- Expanding	<ul style="list-style-type: none"> specific and some technical language of the content areas a variety of sentence lengths of varying linguistic complexity in oral discourse or multiple, related sentences or paragraphs oral or written language with minimal phonological, syntactic or semantic errors that do not impede the overall meaning of the communication when presented with oral or written connected discourse with sensory, graphic or interactive support
3- Developing	<ul style="list-style-type: none"> general and some specific language of the content areas expanded sentences in oral interaction or written paragraphs oral or written language with phonological, syntactic or semantic errors that may impede the communication, but retain much of its meaning, when presented with oral or written, narrative or expository descriptions with sensory, graphic or interactive support
2- Beginning	<ul style="list-style-type: none"> general language related to the content areas phrases or short sentences oral or written language with phonological, syntactic, or semantic errors that often impede the meaning of the communication when presented with one- to multiple-step commands, directions, questions, or a series of statements with sensory, graphic or interactive support
1- Entering	<ul style="list-style-type: none"> pictorial or graphic representation of the language of the content areas words, phrases or chunks of language when presented with one-step commands, directions, WH-, choice or yes/no questions, or statements with sensory, graphic or interactive support oral language with phonological, syntactic, or semantic errors that often impede meaning when presented with basic oral commands, direct questions, or simple statements with sensory, graphic or interactive support

Figure 3. Performance Level Definitions from the 2007 WIDA ELD Standards

The original WIDA Writing Rubric was not intended as an analytic scale. Although raters awarded scores in all three categories, they followed strict guidelines. They were first to decide which holistic level the paper belonged to (e.g. 3); then, if they felt there was a strength or weakness in one category, they could award the next higher or next lower level in that

category, with a limit of one strength or weakness. Such guidelines were intended to provide “diagnostic” information, since the three separate category scores for each writing task were provided on the student’s score report. However, psychometrically speaking, the three category scores were summed, which made, in effect, an 18-point scale (Table 2). (Note: A score of 0 was for blank papers, totally illegible responses, or responses in a language other than English. A score of 0-1-0 was reserved for responses containing only English words copied from the prompt. It was intended to give a nod to vocabulary usage.)

Table 2
Conversion of Raters Scores Using Original WIDA Writing Rubric into Numeric Scores Used for Psychometric Purposes

Possible Rater Score Across Three Categories	“Psychometric” Score used for Producing Scale Scores
0	0
0-1-0	1
1-1-1	3
2-1-1, 1-2-1, 1-1-2	4
1-2-2, 2-1-2, 2-2-1	5
2-2-2	6
3-2-2, etc.	7
2-3-3, etc.	8
3-3-3	9
4-3-3, etc.	10
3-4-4, etc.	11
4-4-4	12
5-4-4, etc.	13
4-5-5, etc.	14
5-5-5	15
6-5-5, etc.	16
6-6-5, etc.	17
6-6-6	18

This approach to scoring the writing tasks was used throughout operational ACCESS, up to ACCESS 2.0.

From Writing Rubrics to Writing Score Scale for ACCESS 2.0

In 2012, WIDA published an amplification of the WIDA ELD Standards. Some of the modifications included the reorganization of the performance level definitions into three categories and the separation of the productive language domains (Speaking and Writing) from the receptive language domains (Listening and Reading). Figure 4 presents the Speaking and

Writing performance level definitions as presented in the 2012 amplification of the WIDA Standards.



WIDA Performance Definitions Speaking and Writing, Grades K-12

At each grade, toward the end of a given level of English language proficiency, and with instructional support, English language learners will produce...

	Discourse Level	Sentence Level	Word/Phrase Level
	Linguistic Complexity	Language Forms and Conventions	Vocabulary Usage
Level 6 – Reaching Language that meets all criteria through Level 5, Bridging			
Level 5 Bridging	<ul style="list-style-type: none"> Multiple, complex sentences Organized, cohesive, and coherent expression of ideas 	<ul style="list-style-type: none"> A variety of grammatical structures matched to purpose and nearly consistent use of conventions, including for effect A broad range of sentence patterns characteristic of particular content areas 	<ul style="list-style-type: none"> Technical and abstract content-area language Words and expressions with precise meaning related to content area topics
Level 4 Expanding	<ul style="list-style-type: none"> Short, expanded, and some complex sentences Organized expression of ideas with emerging cohesion 	<ul style="list-style-type: none"> A variety of grammatical structures and generally consistent use of conventions Sentence patterns characteristic of particular content areas 	<ul style="list-style-type: none"> Specific and some technical content-area language Words and expressions with multiple meanings or common collocations and idioms across content areas
Level 3 Developing	<ul style="list-style-type: none"> Short and some expanded sentences with emerging complexity Expanded expression of one idea or emerging expression of multiple related ideas 	<ul style="list-style-type: none"> Repetitive grammatical structures with occasional variation and emerging use of conventions Sentence patterns across content areas 	<ul style="list-style-type: none"> Specific content words and expressions (including content-specific cognates) Words or expressions related to content areas
Level 2 Emerging	<ul style="list-style-type: none"> Phrases or short sentences Emerging expression of ideas 	<ul style="list-style-type: none"> Formulaic grammatical structures and variable use of conventions Repetitive phrasal and sentence patterns across content areas 	<ul style="list-style-type: none"> General content words and expressions (including common cognates) Social and instructional words and expressions across content areas
Level 1 Entering	<ul style="list-style-type: none"> Words, phrases, or chunks of language Single words used to represent ideas 	<ul style="list-style-type: none"> Simple grammatical constructions (e.g., commands, Wh- questions, declaratives) Phrasal patterns associated with common social and instructional situations 	<ul style="list-style-type: none"> General content-related words Everyday social and instructional words and familiar expressions

...within sociocultural contexts for language use.

Figure 2. Performance Level Definitions for Speaking and Writing from the 2012 Amplification of the WIDA ELD Standards

With the new generation of ACCESS, the WIDA Writing Rubric was updated to reflect the 2012 amplified performance level definitions. Over several years, WIDA and CAL staff worked together to develop the Revised Writing Rubric of the WIDA Consortium (see Figure 5), examining papers that had been previously rated at a whole score point (e.g., 3-3-3) according to the original WIDA Writing Rubric, and ensuring that such papers would receive that same whole score point under the revised rubric (e.g., 3).

Revised Writing Rubric of the WIDA™ Consortium Grades 1-12 (08/29/14)			
	Discourse Level	Sentence Level	Word/Phrase Level
	Linguistic Complexity	Language Forms and Conventions	Vocabulary Usage
Level 6 Reaching	<p>Text is fully comprehensible and appropriate to purpose, situation, and audience; comparable to the writing of English proficient students meeting college- and career-readiness standards; it includes:</p> <ul style="list-style-type: none"> extended connected text (single or multiple paragraphs) that is organized and shows tight cohesion in the precise expression of ideas clear evidence of consistency in conveying an appropriate perspective, register, and genre 	<ul style="list-style-type: none"> a full range of sentence patterns and grammatical structures matched to content area topics consistent use of appropriate conventions to convey meaning, including for effect 	<ul style="list-style-type: none"> consistent usage of just the right word or expression in just the right context related to content area topics facility with precise vocabulary usage in general, specific, or technical language
Level 5 Bridging	<p>Text is comprehensible and generally related to purpose, situation, and audience; generally comparable to the writing of English proficient peers; it includes:</p> <ul style="list-style-type: none"> extended connected text (single or multiple paragraphs) that is organized and shows a cohesive and coherent expression of ideas clear evidence of conveying an appropriate perspective, register, and genre 	<ul style="list-style-type: none"> a broad range of sentence patterns and grammatical structures matched to the content area topic nearly consistent use of appropriate conventions to convey meaning, including for effect 	<ul style="list-style-type: none"> usage of technical and abstract content-area words and expressions as appropriate usage of words and expressions with precise meaning related to content area topics as appropriate vocabulary usage that fulfills the writing purpose
Level 4 Expanding	<p>Text is generally comprehensible and related to purpose; it includes:</p> <ul style="list-style-type: none"> connected text (sentences or paragraphs) that shows an organized expression of ideas with emerging cohesion some evidence of conveying an appropriate perspective, register, and genre 	<ul style="list-style-type: none"> a range of sentence patterns and grammatical structures characteristic of the content area generally consistent use of appropriate conventions to convey meaning 	<ul style="list-style-type: none"> usage of specific and some technical content-area words and expressions as appropriate usage of words and expressions with multiple meanings or common collocations and idioms across content areas as appropriate vocabulary usage that generally fulfills the writing purpose
Level 3 Developing	<p>Original text is generally comprehensible (though comprehensibility may from time to time be compromised in more complex original text); it includes:</p> <ul style="list-style-type: none"> text that shows developing organization in the expression of an expanded idea or multiple related ideas evidence of a developing sense of perspective, register, and genre 	<ul style="list-style-type: none"> a developing range of sentence patterns and grammatical structures common to content areas developing use of conventions to convey meaning 	<ul style="list-style-type: none"> usage of some specific content words and expressions as appropriate usage of words or expressions used frequently in content areas, as appropriate vocabulary usage that attempts to fulfill the writing purpose
Level 2 Emerging	<p>Some original text and text adapted from model or source text is generally comprehensible (though comprehensibility may often be compromised in attempts at more complex original text); it includes:</p> <ul style="list-style-type: none"> text that shows emerging expression of an idea or ideas and may demonstrate some attempt at organization some amount of text that may be copied or adapted 	<ul style="list-style-type: none"> repetitive sentence and phrasal patterns and formulaic grammatical structures used in social and instructional situations or across content areas variable use of conventions 	<ul style="list-style-type: none"> usage of general content words and expressions usage of social and instructional words and expressions across content areas possible usage of general vocabulary where more specific language is needed
Level 1 Entering	<p>Text that is copied or adapted from model or source text is generally comprehensible (though comprehensibility may be significantly compromised in original text); it includes:</p> <ul style="list-style-type: none"> language that represents an idea or ideas varying amounts of text that may be copied adapted text that may contain some original language 	<ul style="list-style-type: none"> words, chunks of language, or simple phrasal patterns associated with common social and instructional situations possible use of some conventions 	<ul style="list-style-type: none"> usage of highest frequency general content-related words usage of everyday social and instructional words and expressions

Figure 3. Revised Writing Rubric of the WIDA Consortium 2014

The revised rubric was piloted in the fall and winter of 2014-15 by raters from operational ACCESS to score responses from the first phase of field testing, conducted in 2013-14. Raters were instructed to award only a whole-score level (e.g., 3), though as with the former rubric, they could award a plus (e.g., 3+) if there was a particular strength in one or two categories, or a minus (e.g., 3-) if there was a particular weakness in one or two categories. Several anchor sets were provided for rater training.

Following a first scoring session, CAL and WIDA staff debriefed with raters who led the field test scoring, and it was determined that there were some misunderstandings about how to apply the revised rubric. A second scoring session immediately followed, in which previously scored papers were re-scored, such that all papers were double scored. Following the second session, scores were sent to CAL and a second debrief with lead raters was held. Based on the data collected, it appeared that the revised rubric was not working as intended because raters thought the rubric was complicated to apply and contained too much text (i.e., there were too many factors to consider). Upon further analysis, it was determined that the revised rubric was too unwieldy and was oriented more toward writer capability than toward the evidence of that capability within the written product.

CAL staff then proposed a simplified rubric that 1) would be easier to use, 2) was directed to citing evidence in the written product, and 3) would focus rater attention to the discourse, sentence, and word/phrase levels, despite a looser relationship to the 2012 performance level definitions than the 2014 revised writing rubric (Figure 5). With WIDA's consent, CAL staff examined the whole-score written responses across grade levels that were used to develop the first rubric (e.g., 1-1-1, 2-2-2, 3-3-3, etc.). The goal was to determine, in each category, what features most clearly distinguished one level from another, and to state this succinctly from a rater perspective. After several rounds of revision, this rubric was delivered to the operational scoring program in preparation of scoring the writing responses collected during Phase 2 of the field test (2014-15), at which time the writing responses collected during Phase 1 would also be re-scored. The revised 2014 rubric came to be internally called the "interpretive rubric," (i.e., intended to help score users understand writing scores on ACCESS 2.0), while the new, simplified rubric came to be labeled the Writing Score Scale and would not be intended for public consumption, since its relationship to the performance level definitions would be less transparent. In addition, when using the new Writing Score Scale, *all* papers between two whole levels would receive a plus (+) score; minus scores would disappear.

When the new Writing Score Scale and its anchor papers (which were predominantly the same anchors as before) were finalized, the lead raters for the field test scoring examined the scale and the anchors very carefully. After some minor revision and appropriate approvals by CAL

and WIDA, they were delivered to the scoring vendor for the scoring of field test writing responses.

The Writing Score Scale used for scoring all papers from the field test is presented in Figure 6.

ACCESS for ELLS 2.0 Writing Scoring Scale (Revised 7/9/2015)	
5+	<p>Score Point 6</p> <p>D: Sophisticated organization of text clearly demonstrating an overall sense of unity throughout, tailored to context (e.g., audience, purpose and situation)</p> <p>S: Purposeful use of a variety of sentence structures that are essentially error-free</p> <p>W: Precise use of vocabulary with just the right word in just the right place</p>
	<p>Score Point 5</p> <p>D: Strong organization of text supporting an overall sense of unity, appropriate to context (e.g., audience, purpose and situation)</p> <p>S: A variety of sentence structures with very few grammatical errors</p> <p>W: A wide range of original vocabulary, used appropriately and with ease</p>
	<p>Score Point 4</p> <p>D: Organized text presenting a clear progression of ideas demonstrating an awareness of audience and purpose</p> <p>S: Complex and some simple sentence structures, containing occasional grammatical errors that don't generally interfere with comprehensibility</p> <p>W: A variety of vocabulary beyond the prompt, generally conveying the intended meaning</p>
3+	<p>Score Point 3</p> <p>D: Text that shows developing organization including the use of elaboration and detail, though the progression of ideas may not always be clear</p> <p>S: Simple and some complex sentence structures, whose meaning may be obscured by noticeable grammatical errors</p> <p>W: Some vocabulary beyond the prompt, although usage is noticeably awkward at times</p>
	<p>Score Point 2</p> <p>D: Text that shows emerging organization of ideas but with heavy dependence on the stimulus and prompt and/or resemblance to a list of simple sentences (which may be linked by simple connectors)</p> <p>S: Simple sentence structures; meaning is frequently obscured by noticeable grammatical errors when attempting beyond simple sentences</p> <p>W: Vocabulary primarily drawn from the stimulus and prompt</p>
1+	<p>Score Point 1</p> <p>D: Minimal text that represents an idea or ideas</p> <p>S: Primarily words, chunks of language and short phrases rather than complete sentences</p> <p>W: Distinguishable English words that are often limited to high frequency words or reformulated expressions from the stimulus and prompt</p>
	<p>0+ <i>Consists only of verbatim copied text that is related to the prompt</i></p>
0	<p><i>Blank or no response; or consist only of text (copied or not) that is completely off-task; or entirely in a language other than English</i></p>

Figure 4. WIDA Writing Scoring Scale Used to Score All Field Test Papers 2015

Because the goal in revising the writing rubric was to ensure that there would be similar scores awarded to papers using the original or the revised rubric, the conceptual relationship between the two can be shown, as in Table 2. The table shows that the scores at the whole score points should be retained, and scores between score points would be collapsed from two categories into one. In addition, it was decided that copying English words (which received a score of 0-1-0

on the original rubric and a score of 0+ on the revised score scale) is not an indication of developing English language proficiency, so no distinction between 0 and 0+ would be made for the psychometric analyses nor for operational ACCESS 2.0 scoring. It should also be noted that the score report would be simplified for ACCESS 2.0, reporting only the Writing score scale without any raw score ratings.

The slight change in the rating to raw score conversion from ACCESS to ACCESS 2.0, however, will require that the step measures associated with the psychometric rating scale for the ACCESS Writing Test be reestablished for the ACCESS 2.0 Writing Test. The plan to do so is described in more detail below.

Table 3
Conceptual Relationship of “Psychometric” Scores between Responses Scored with the Original WIDA Writing Rubric and the New WIDA Writing Scoring Scale

Original WIDA Writing Rubric (2006-2015)		WIDA Writing Scoring Scale (Field Test for ACCESS 2.0)	
Possible Rater Score Across Three Categories	“Psychometric” Score for Producing Scale Scores	Rater Scores	“Psychometric” Score for Producing Scale Scores
0	0	0	0
0-1-0	1	0+	0
1-1-1	3	1	1
2-1-1, 1-2-1, 1-1-2	4	1+	2
1-2-2, 2-1-2, 2-2-1	5	2	3
2-2-2	6	2+	4
3-2-2, etc.	7	3	5
2-3-3, etc.	8	3+	6
3-3-3	9	4	7
4-3-3, etc.	10	4+	8
3-4-4, etc.	11	5	9
4-4-4	12	5+	10
5-4-4, etc.	13	6	11
4-5-5, etc.	14		
5-5-5	15		
6-5-5, etc.	16		
6-6-5, etc.	17		
6-6-6	18		

Maintaining the Underlying Psychometric Writing Scale

The ACCESS 2.0 Writing field test¹ data are used to link ACCESS 2.0 operational tasks to ACCESS tasks in order to maintain the ACCESS Writing score scale. The ACCESS 2.0 Writing field test was designed with the aim of collecting a sample of student performances on both the ACCESS 2.0 field test and the operational ACCESS test such that a common-person design can be used to establish the link between performances on the two tests. The ACCESS 2.0 Writing field test tasks were administered to samples of students within two weeks of taking operational ACCESS. The target sample size was 350 students per field test task. All ACCESS 2.0 Writing field test tasks were adapted from Writing tasks previously used on ACCESS. For logistical reasons, however, the ACCESS 2.0 Writing field test was administered under conditions that deviate from the operational ACCESS 2.0 Writing Test administration in two ways. First, during the ACCESS 2.0 field test, Grades 4-5 students wrote their responses on paper, while operational ACCESS 2.0 will allow Grades 4-5 students to either write by hand or keyboard their responses. Also, during the ACCESS 2.0 field test, Grades 6-12 Tier A students wrote their responses by hand while operationally, all Grades 6-12 students will keyboard their responses unless writing by hand is provided to individual students as an accommodation².

The second difference between the ACCESS 2.0 Writing field test and operational ACCESS 2.0 Writing is the time allotment. Operationally, a minimum of one hour is needed for students to take all three Writing tasks, but the ACCESS 2.0 Writing field test was limited to 45 minutes. Consequently, students either took two short tasks or one long task during field test administration, as opposed to three tasks in the operational ACCESS and ACCESS 2.0 administrations. Because of these deviations in administration conditions, the parameters derived from the field test should be considered preliminary. As described below, a verification study will be conducted using the ACCESS 2.0 operational data to revise these parameters.

Figure 7, located in the Appendix at the end of this read-ahead, illustrates the four-step plan to link field test and operational Writing scores via a common-person design: (I) an initial common-person linking, (II) Outlier analysis, (III) final common-person linking, and (IV) a verification procedure. The steps are described in greater detail below.

¹ ACCESS 2.0 Series 400 Online Writing field test was conducted in both 2013-2014 and 2014-2015 academic year.

² Concerns about the comparability between handwritten vs. keyboarded responses are addressed in Read-ahead 2.5.

Step I. Initial Common-Person Linking

Obtain students' ability parameters on the ACCESS Writing Test.

Fix the students' ability parameters while estimating the step parameters and task difficulty parameters for the ACCESS 2.0 Writing Test tasks. The calibration will be conducted across grade-level clusters. Because the same rubric is used to score all writing samples across all clusters, a single rating scale can be used. Through this concurrent calibration procedure, the step parameters and task difficulty parameters of the ACCESS 2.0 Writing field test will be placed on the ACCESS Writing score scale. The fit of the rating scale to the data will be evaluated in Step C and adjustments will be made to the rating scale as needed.

Examine the infit and outfit statistics of the step parameters across all grade-level clusters. For each cluster, the step parameters will be anchored at the values computed from Step B and the infit and outfit statistics of the step parameters will be evaluated to determine whether the rating scale fit the data. If the infit and outfit statistics indicate that the rating scale does not fit the data for a particular grade-level cluster, the step parameters may be re-estimated for that cluster.

Step II: Outlier Analysis

The rationale for conducting an outlier analysis was discussed at the 2014 TAC meeting and reviewed in read-ahead 2.1. Because each ACCESS 2.0 field test form consisted of only one or two tasks, it is not feasible to use exactly the same methodology that was used to identify outliers in the ACCESS 2.0 Listening and Reading field test analysis. Instead of identifying outliers prior to conducting the linking analysis, outlier analysis will be conducted after the initial linking analysis is completed. Procedurally, after the initial common-person linking is conducted in Step I.B, the person measures estimated based on ACCESS 2.0 task performances will be compared with measures estimated based on performances on operational ACCESS tasks. Then, the scatter plot procedure in the Rasch-based software Winsteps is used to identify students that exhibit statistically significant differences (as evidenced by their person measures) between two administrations. Students with t-statistics greater than ± 2 will be classified as outliers and removed from the final linking analysis.

Step III. Revise Task Difficulty Parameters

After the outliers are removed, a second round of common-person linking will be conducted to revise the task difficulty parameters while the step measures will be fixed to the values derived from Step I.

Step IV. Verification Study

For score reporting for ACCESS 2.0, task difficulty parameters will be refined through a verification procedure using operational ACCESS 2.0 Writing data collected during the early testing window, which is analogous to what is currently done for the operational ACCESS administration. The ACCESS 2.0 Writing Test task difficulty parameters will be initially anchored to the values derived from the ACCESS 2.0 field test administration analyses, and displacement statistics will be evaluated to determine whether the parameters need to be re-estimated based on the equating sample data. The final task difficulty parameters derived in this step will be used to score students for the operational ACCESS 2.0 administration.

Questions for the TAC:

- Question 1: Do the methods used to maintain the Writing score scale appear sound?
- Question 2: What additional analyses might the TAC recommend to ensure that the linking is conducted appropriately and that the Writing reporting scale is maintained?

**APPENDIX
Common-Person Design**

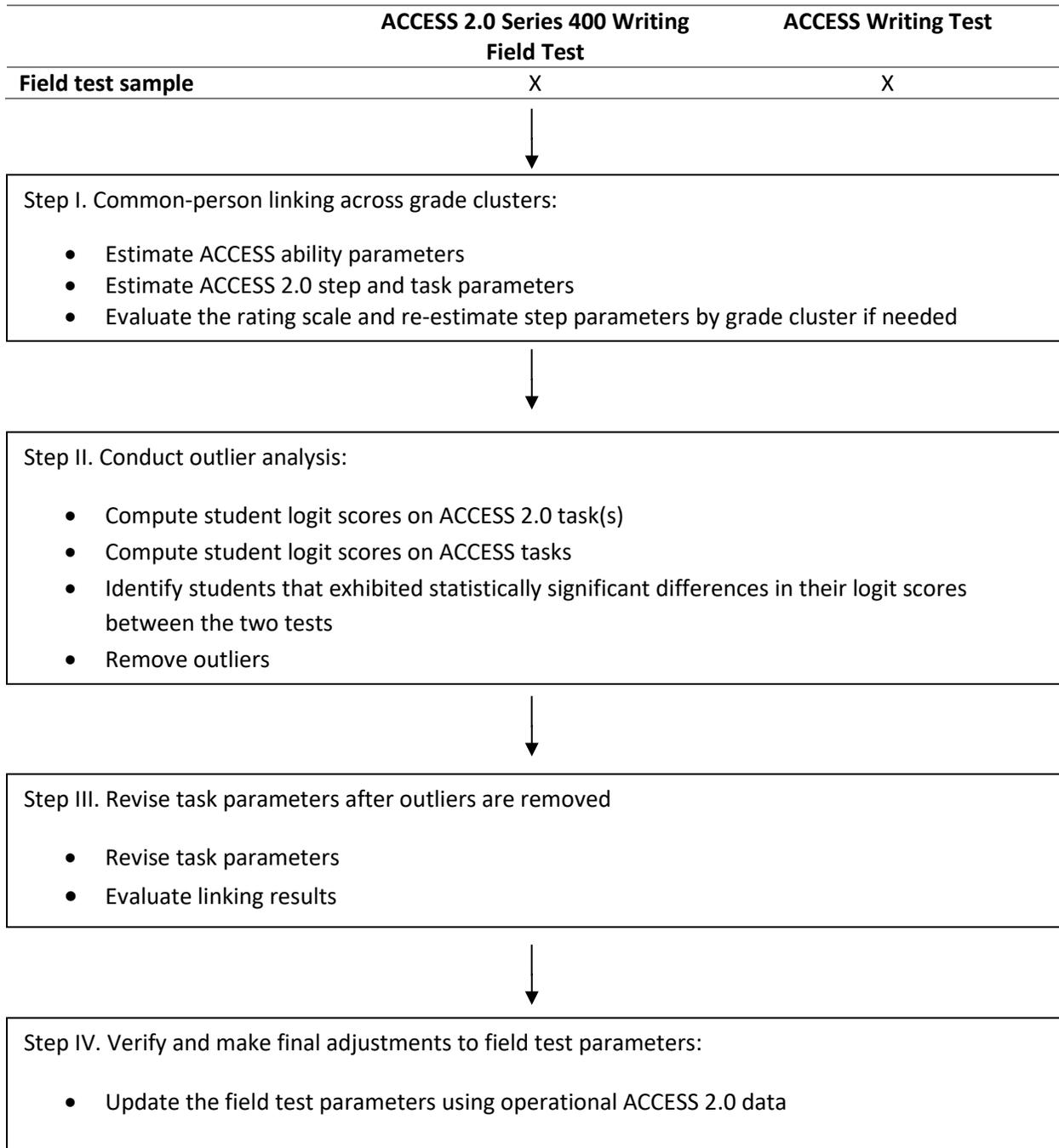


Figure 7. Proposed method to maintain the ACCESS Writing Test score scale