



**A Seal of Biliteracy for All: Expanding Assessment  
Options Toward Access & Equity for Learners of Less  
Commonly Tested Languages**

**Linda Egnatz**  
Executive Director  
Global Seal of Biliteracy

**Victor D. O. Santos**  
Director of Assessment  
and Research  
Avant Assessment

**July 2021**

## TABLE OF CONTENTS

<b>Introduction</b> .....	<b>2</b>
1.1 An Overview of State Seals of Biliteracy .....	3
1.2 Biliteracy Defined .....	4
1.3 Common Test Criteria for State Seals of Biliteracy (4-skill measures) .....	5
1.4 Access & Equity Problem: Limited 4-skill Tests for LCTLs .....	5
1.5 The Impact of Limited 4-Skill Assessments .....	6
1.6 The (Unintended) Equity Gaps in State Seal Programs .....	7
1.7 An Overview of the Global Seal of Biliteracy .....	8
<b>Literature review</b> .....	<b>9</b>
2.1 The Importance of Testing as many Skills as Possible .....	9
2.2 Correlation Among Language Skills and Relationship Between Receptive and Productive Knowledge .....	9
2.2.1 Reading and Writing Skills .....	10
2.2.2 Listening and Speaking Skills .....	10
<b>The Present Study</b> .....	<b>11</b>
3.1 Research Questions .....	11
3.2 Materials .....	12
3.3 Participants .....	12
3.4 Results .....	13
<b>Discussion</b> .....	<b>25</b>
<b>References</b> .....	<b>27</b>

### **Suggested Citation**

Egnatz, L. & D.O. Santos, V. (2021). *A Seal of Biliteracy for All: Expanding Assessment Options Toward Access & Equity for Learners of Less Commonly Tested Languages*

# 1. INTRODUCTION

The year 2021 marks the 10<sup>th</sup> anniversary of the Seal of Biliteracy (SoBL). Begun in California in 2011, the goal of this state-by-state policy initiative is to “recognize a student who has attained proficiency in English and one or more other world languages by high school graduation” (ACTFL, NABE, NCSSEFL, & TESOL, 2015, p. 2). The seal, which is affixed to the student’s high school diploma or transcript, represents a cultural shift by both political and educational institutions that, for the first time, celebrate the value of being bilingual. Following the grassroots success of California in 2011, 42 states and the District of Columbia now offer a Seal, Certificate or Endorsement of Biliteracy program (Figure 1).

Adoption Year	State(s)
2011	California
2013	Texas, New York, Illinois
2014	New Mexico, Washington, Minnesota, Louisiana, District of Columbia
2015	North Carolina, Virginia, Indiana, Nevada, Hawai’i, Wisconsin, Utah
2016	New Jersey, Oregon, Florida, Maryland, Georgia, Arizona, Kansas, Rhode Island
2017	Ohio, Colorado, Connecticut, Delaware, Missouri, Massachusetts
2018	Michigan, Tennessee, Iowa, Arkansas, South Carolina, Maine
2019	North Dakota, Mississippi
2020	Nebraska, Idaho, Oklahoma, New Hampshire
2021	West Virginia

Figure 1: States with bilingual recognition programs and date of adoption.

The SoBL movement, which has quickly been adopted across the country, was created by [Californians Together](#) partly to change common paradigms about English Learners (ELs) and to recognize their languages as assets rather than deficits. Formally awarding a Seal documents language skills and may provide bilinguals with benefits such as competency-based high school language credit for a language not taught at their school, and in some cases advanced placement or college credit. For heritage language learners, the SoBL becomes a valuable incentive to sustain and grow their home language(s) and can lead to greater career opportunities for those learners. As an added benefit, it also serves to protect the rich cultural heritage and identity of America’s languages.

In an effort to provide guidelines to the states and some standardization, [ACTFL](#) (formerly the American Council on the Teaching of Foreign Languages), [Teachers of English to Speakers of Other Languages](#) (TESOL), the [National Association for Bilingual Education](#) (NABE), and the [National Council of State Supervisors for Languages](#) (NCSSEFL) joined to create and publish the *2015 State Guidelines for Implementing the Seal of Biliteracy* (ACTFL et al., 2015). The document recommends that:

- (a) both native and non-native speakers of English provide comparable evidence of English Proficiency, as determined by the state guidelines;
- (b) native and non-native users of a language other than English provide evidence of proficiency;
- (c) the minimum target level should be Intermediate-Mid based on the ACTFL Proficiency Guidelines;

(d) special allowances should be made due to the unique characteristics of certain languages (Classical Languages, American Sign Language (ASL), and Native American Languages).

(e) to the extent possible, all language skills available for the language in question should be assessed.

## 1.1. An Overview of State Seals of Biliteracy

The various state Seals of Biliteracy share a common goal, but state SoBL rules, procedures, and rewards vary greatly from state to state and often among districts within the same state. These differences begin with how each state adopted its program. Many of the early adopting states began grassroots district-level programs prior to implementation of statewide policies. Some states have enacted their SoBL by passing state legislation, some through resolution by the state board of education and still others by providing general program guidelines to local district administrators. Each state's program is designed to recognize the biliteracy of their graduating seniors, but the states have chosen a variety of names among which are Seal of Biliteracy, Certificate of Multilingualism, and Global Languages Endorsement. Arizona, for example, offers a Seal of Bilingualism-Biliteracy in order to include its Native American languages that may not have a written form. The SoBL is generally awarded to high school seniors who meet the state's established criteria, and four states (Hawai'i, Illinois, Maine, Minnesota) provide college credit at state university systems to their recipients. SoBL recipients may also qualify for competency-based world language credits for languages not taught in their school. In a Washington case study, 21% of EL (English Learner) students needed the language credits earned by testing for the Washington SoBL in order to graduate and 10% needed the world language credits to apply to a four-year college program (Greenberg Motamedi & Jaffery, 2015).

State SoBL programs also vary greatly in the criteria required for students to achieve the award (Davin & Heineke, 2017). States, per the guidelines, require evidence of proficiency in both English and another language. Despite the recommendation that all languages should be assessed in all language skills available, the requirements by state for English language proficiency are diverse and include Grade Point Average (GPA), meeting graduation requirements, or passing a standardized or end-of-course exam. It is noteworthy that none of these "demonstrations of English language proficiency" include measures of Listening or Speaking skills. ELs, who make up an estimated 10% of the K-12 student population in the USA (National Center for Education Statistics, 2021), often have to meet more rigorous English proficiency requirements than English-dominant students. This may include passing a test such as the ACCESS test, created by the 40-state WIDA Consortium and which includes listening and speaking as well as demonstrating heritage language proficiency. This and other equity issues, such as access to a test of a home language, create barriers to the SoBL for many ELs (Heineke, Davin & Bedford, 2018).

For languages other than English (LOTEs), most states require that a proficiency level be demonstrated via testing of a designated level of language proficiency based on the ACTFL Proficiency Guidelines (ACTFL, 2012). The Guidelines identify five major levels of proficiency (Novice, Intermediate, Advanced, Superior, and Distinguished) and three sub-levels (Low, Mid, High), with the sublevels applying only to the Novice, Intermediate and Advanced levels. There are however a few states (California, Florida, Louisiana, North Carolina, and Texas) where students can demonstrate proficiency via *seat time* by achieving a particular GPA in world language coursework (see Figure 2). Several states offer more than

one level of SoBL, as recommended by the 2015 guidelines, especially as a way to recognize heritage and dual-language learners. Most states provide schools with a list of approved qualifying assessments for LOTEs, and some states offer alternative methods such as a portfolio option for students of less commonly taught languages (LCTLs) or for those with disabilities. For most state SoBL programs, however, the rules require that proficiency in LOTEs be demonstrated in all four (4) skills, namely the two receptive skills of Reading and Listening, as well as the productive skills of Writing and Speaking, as recommended by ACTFL et al. (2015).

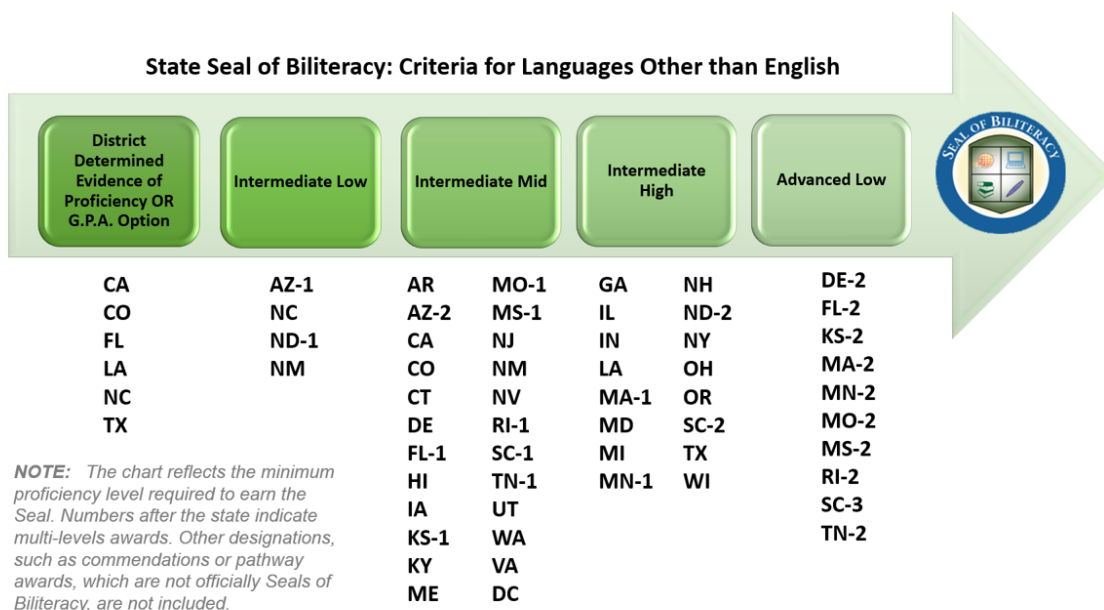


Figure 2. Minimum language proficiency level required for the SoBL across different states.

## 1.2. Biliteracy Defined

Almost all state SoBL programs use the term “biliteracy” rather than “bilingualism.” The language of the legislation often includes phrases such as academic excellence, high level of proficiency, and even rigor. Bill SB1221 Section 1 of the Illinois Senate (Illinois General Assembly, 2013) states that “it is the intent of the legislature to promote linguistic proficiency and cultural literacy in one or more languages in addition to English and to provide recognition of the attainment of those needed and important skills through the establishment of the State Seal of Biliteracy.” The inclusion of literacy, generally defined as the ability to read and write, compels the SoBL criteria to include evidence of each. In simple terms, a bilingual person can speak two languages whereas a biliterate person can read and write in two languages. A person who is biliterate is also considered bilingual, but a person who is bilingual is not necessarily biliterate.

### 1.3. Common test criteria for State Seals of Biliteracy (4-Skill measures)

The three 4-skill tests most commonly listed as approved assessments for languages other than English to earn the SoBL are included in Figure 3 below along with the number of languages each test offers and the number of SoBL states that accept the test. Some states also accept 4-skill individual language tests based on the Common European Framework of Reference for Languages scale (CEFR).

Language Assessment	# of Languages Offered	# of States Accepting Test (Some new adoptions have not yet posted criteria)
Avant Standards-based Measurement of Proficiency (Avant STAMP 4S)	14	40
ACTFL Assessment of Performance toward Proficiency in Languages (AAPPL) Measure	13	40
Advanced Placement (AP®) College Board	7	38

Figure 3. Number of languages offered and the number of states that approve the three most commonly accepted four-skill assessments for the SoBL.

### 1.4. Access & Equity problem: Limited 4-skill tests for LCTLs.

The challenge for users of less commonly taught or less commonly tested languages (LCTLs) is clear. The report entitled “*Our Nation’s English Learners*” (U.S. Department of Education, 2018) notes that in the 2014-15 school year, ELs in U.S. public schools spoke over 400 languages. The most commonly spoken non-English language was Spanish, followed by Arabic, Chinese and Vietnamese. In Pennsylvania, for example, more than 225 unique languages were reported. A visual of U.S. linguistic diversity is shown in the map below of each state’s most spoken language excluding English and Spanish (Kiersz, Luce, & Hoff, 2020). Sixteen of these languages (Vietnamese, Tagalog, Somali, Dakota, Lakota, Nakota, Sioux, Hmong, Nepali, Navajo, Ilocano, Aleut, Eskimo, Haitian Creole, Pennsylvania Dutch, and Gujarati) are not tested by any of the three major four-skill tests listed in Figure 3. Of the total 22 languages depicted in Figure 4, the AP® test is only available in three, whereas the AAPPL test is offered in five and the Avant STAMP 4S test, which at the moment of this writing covers the most languages among the three testing options, is available in six of them.



Given the issues outlined above regarding suboptimal alternative assessment of languages that do not benefit from having a standardized 4-skill test available, a growing number of states, most recently Massachusetts, have begun to accept as an alternative, a standardized, validated 2-skill test that measures the productive skills of speaking and writing. Washington state has aggressively worked, often with third party agencies, to identify and fund tests in order to provide their bilingual students with both a SoBL and [competency based World Language Credits](#). Unfortunately, however, many school districts tasked with locating alternative solutions may determine it to be too costly and labor intensive and decide not to offer the program. In spite of the large numbers of ELs and the variety of languages they represent nationally, only a handful of states awarded the SoBL in 25 or more languages in the academic year 2018-2019 (Black, Chou, & Hancock, 2020).

### States Awarding Seals in 25 or More Languages 2018-19

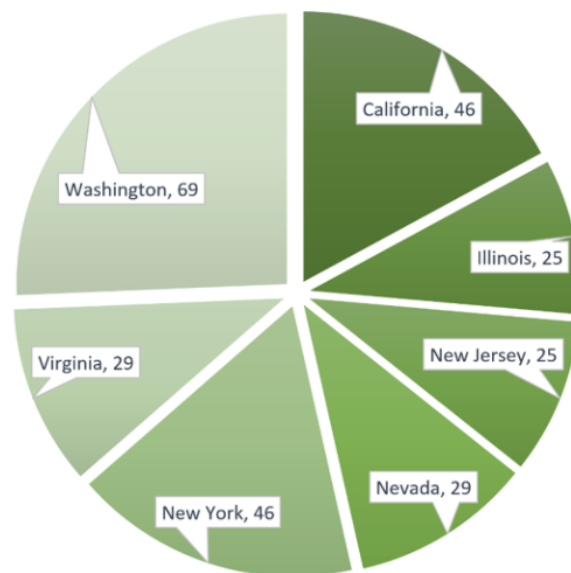


Figure 5. States that Awarded the Seal in 25 or More Languages in 2018-2019 (Black et al., 2020).

### 1.6. The (Unintended) Equity Gaps in State Seal Programs

When states pass their Seal of Bilingualism into law, the legislation often only includes public schools because states have no statutory authority to regulate or monitor private schools. The unfortunate result is that in many states, students enrolled in non-publicly-funded schools such as private, parochial, charter, and home schools are excluded from the SoBL opportunity, as are bilinguals in non-participating public schools. Because the state SoBL is not mandated and is not publicly funded, many schools that qualify do not participate. In the 2018-19 year, of 16 states reporting, only 21% of public schools opted to award the SoBL (Black et al., 2020). Students who have learned a language outside of a traditional classroom may also be disenfranchised unless the school system provides for testing of languages not taught in the school or accepts test scores from external sources such as a community-based heritage language school. Another group not included in the state SoBL, but who could greatly benefit from a way to document their bilingualism, are students enrolled in Higher Education.



## 1.7. An Overview of the Global Seal of Bilingualism

To address the frustration of the bilinguals who cannot participate in the state SoBL and to fill the opportunity gaps left by the state SoBL programs, the [Global Seal of Bilingualism](#) (GSoB) initiative was launched in 2018. The GSoB provides a free unique serial-numbered language credential to anyone, anywhere, of any age who can demonstrate their language skills via external, proctored testing in over 120 languages. Since its inception, the GSoB has provided language credentials to thousands of bilinguals, 33% of whom were identified as current or former EL students. The Global Seal, which does not have age or grade restrictions, has been awarded to students at the college, high school and middle school levels as well as to adults in North America, Asia and Europe. The GSoB encourages learners to “level up” and offers a pathway of language credentials termed “Functional Fluency,” “Working Fluency,” and “Professional Fluency” on both the ACTFL (Intermediate-Mid, Advanced-Low, Advanced-High, respectively) and CEFR (B1, B2, C1, respectively) proficiency scales. Criteria for earning the serial-numbered GSoB meta-credential is the same for everyone, regardless of their age or languages.

The Global Seal of Bilingualism’s independent [Board of Advisors](#) is responsible for approving tests used to determine a candidate’s level of language proficiency. Guiding Principles were created with the primary goal of expanding access to credentialing. Where possible, the GSoB determined that approved tests should assess all four (4) basic language skills: Listening, Speaking, Reading and Writing. From the beginning, exceptions were made for LCTLs (less commonly taught languages) where four-skill tests are not available or are significantly cost prohibitive. To provide access to language credentials for these bilinguals, a test that measures the two (2) productive skills of speaking and writing were approved, a decision also made by states such as Massachusetts and Michigan for the SoBL, as previously noted.

By allowing for these accommodations, the Global Seal was able to greatly expand language credentialing to ELs, Heritage Language Learners (HLLs) and those who acquired their skills outside of participating public school classrooms, without having to resort to non-standardized, alternative methods of measuring language proficiency. For bilinguals not supported by their schools, the GSoB offers an individual application to provide access for all language learners.

It is important to note that the Global Seal firmly believes that, whenever a standardized and validated 4-skill assessment is available in a given language, that should take priority over an equally standardized and validated 2-skill assessment of that language. It is also a belief held by the Global Seal that the benefits of accepting a standardized and validated 2-skill assessment such as the Avant STAMP Writing and Speaking ([Avant STAMP WS](#)) test (formerly known as WorldSpeak) or the LTI/WPT combination for a language that currently lacks a defensible four-skill assessment far outweigh the risks of employing less rigorous assessments or no assessment at all.

## 2. LITERATURE REVIEW

### 2.1. The Importance of Testing as Many Skills as Possible

As language teachers and language proficiency researchers have long known, each of the four skills of Reading, Writing, Listening, and Speaking provide different pieces of evidence towards one's general proficiency in a language. In addition, it is quite defensible to say that the best and most desirable way to test a test-taker's proficiency in a specific skill such as Speaking is by asking them to *speak* instead of asking them to read, write, or listen. Therefore, for a context such as the state Seal of Biliteracy or the Global Seal of Biliteracy, in which we are indeed interested in gathering accurate information about a test-taker's proficiency in each of the four skills, a well-developed and validated language proficiency assessment in which all four skills are assessed should take priority and be preferred over another equally well-developed and validated proficiency assessment in that same language that does not assess all four skills. If there is further evidence that can be collected about each of the skills, test users should strive to collect that information in order to make a more informed decision about a test-taker's overall or skill-specific proficiency. After all, the decision to use one specific skill as a proxy (i.e., indirect or surrogate measure) of another *could* lead to some test-takers being disproportionately misclassified, especially those whose language learning trajectory has caused their proficiency level in each of the skills to be unbalanced (Powers, 2013).

That being said, we also know that the four language skills correlate with one another, with several types of knowledge being relevant to all of them (knowledge of vocabulary/semantics, grammar/syntax, pragmatics, morphology/word formation, and phonology). As Berninger (2000) notes, the four language skills develop in "overlapping and parallel waves, rather than in discrete, sequential stages (p. 66).

It is also a fact that each language skill provides some level of information about the others, thus correlating with and being interrelated with one another (Schoonen, 2019), although the exact amount of information provided depends on the specific assessment at hand and the reliability of its scores, how each of the four skills is measured, the learning trajectory and profile of the test-takers, and the specific language at hand (Hulstijn, 2015). As Powers (2013) notes, "measurement in non-target domains can be useful in augmenting the prediction of both language development and everyday performance in other domains."

### 2.2. Correlation among Language Skills and Relationship Between Receptive and Productive Knowledge

Laufer, Elder, Hill, and Congdon (2004) argue that there is an empirical implicational hierarchy among different types of knowledge about a word, and that if a test-taker is able to actively recall a word in the L2 (i.e., produce that word without assistance), they should also be able to

recognize it and to passively recall it (i.e., understand its meaning in a passage or identify its meaning from a selected number of options).

The same thinking process can be hypothesized to apply to the receptive skills of Reading and Listening versus the respective productive skills of Writing and Speaking. If a test-taker is able to *produce* language at a certain level, there is good reason and rationale to believe that they should also be able to *understand* language at that same level. After all, we educators have often come across students who are able to understand a certain language to a good extent without being able to produce much, but the reverse is much more rare. Richards (2015) goes as far to say that “all language users have greater receptive competence (language they can understand) than productive competence (language they can produce).

### **2.2.1. Reading and Writing Skills**

As Schoonen (2019) notes “we can expect the same building blocks or constituent components to play a role in the cognitive processes of reading and writing.” Reading and Writing proficiency share a common number of features, including grapheme to word mapping, knowledge of spelling, mapping of word form to meaning, mapping of the graphical representation of a word to its phonetic representation, grammatical knowledge, and pragmatic knowledge.

In a study of 130 male and female Iranian TOEFL iBT test-takers from a variety of mother tongues and ethnic backgrounds, a strong correlation of 0.62 was detected between their Reading scores and their Writing scores (Pearson, 2018). The correlation between their Writing score and their total/overall test score was 0.85, which provides some support to the idea that Writing scores can also be strongly correlated with overall proficiency scores.

### **2.2.2. Listening and Speaking Skills**

In a study of 198 International Teaching Assistants (ITAs) at Temple University (Pennsylvania) who had taken both the Listening and the Speaking sections of the TOEFL iBT test, Wagner (2016) detected a strong correlation of 0.60 between their Listening and Speaking scores. The researcher also detected that their average Listening score was higher than their average Speaking score (24.04 vs 21.74). Pearson (2018) detected a correlation of 0.71 between TOEFL iBT Speaking scores and total/overall score on the test, which once again provides some evidence that that Speaking scores can be strongly correlated with overall proficiency scores.

### 3. The Present Study

In this study, we aim to evaluate the plausibility of using language proficiency scores in the productive skills of Writing and Speaking in order to award State or Global Seals of Biliteracy *in languages that do not currently benefit from the availability of a standardized and validated four-skill assessment in that language*. A case in point can be observed in the languages that Avant Assessment offers through its STAMP WS test, a test solely of productive skills (Writing and Speaking) and built to the same ACTFL guidelines and test development standards as its flagship, 4-skill STAMP 4S test. Languages in which the Avant WS test is available includes Czech, Filipino (Tagalog), Haitian-Creole, Hawaiian, Hmong, Ilocano, Tamil, Turkish, Urdu, Vietnamese, and several others.

The basic assumption underlying the plausibility of using Writing scores as a *proxy* for test-takers' *minimum* expected Reading scores and of using test-takers' Speaking scores as a *proxy* for their *minimum* expected Listening scores is the idea that, as defended by Richards (2015), receptive proficiency will always be at least as high as its productive counterpart. In other words, a given test-taker's Reading proficiency will always be at least as high as their Writing proficiency, and their Listening proficiency will always be as high as their Speaking proficiency.

Fortunately, Avant Assessment has language proficiency results across all four skills for hundreds of thousands of test-takers across many different languages who take the Avant STAMP 4S test each year, being therefore in a position to evaluate the plausibility of the assumption above.

We must note here that the results presented below are only valid and applicable to data pertaining to the Avant STAMP test (both the Avant STAMP 4S and the Avant STAMP WS) since results can be and often are highly dependent on the test development approach and standards employed by each testing company. While similar results may hold for other language proficiency assessments, it is up to the developer of those assessments to show that their test data also supports the assumption.

#### 3.1 Research Questions

We aim to answer the following research questions in the present study:

**Research Question 1:** What is the correlation between Writing - Reading scores and between Listening - Speaking scores on the Avant STAMP 4S test, across various languages?

**Research Question 2:** Could a test-taker's Writing score be reliably used as indirect evidence (*i.e.*, *proxy*) of the *minimum* Reading score the same test-taker would achieve for the purposes of the State or Global Seals of Biliteracy, across various languages?

**Research Question 3:** Could a test-taker's Speaking score be reliably used as indirect evidence (*i.e.*, proxy) of the *minimum* Listening score the same test-taker would achieve for the purposes of the State or Global Seals of Biliteracy across various languages?

### 3.2 Materials

In order to answer the three research questions above, the Avant STAMP 4S language proficiency assessment, developed by Avant Assessment, was selected as the test of choice. Seven Avant STAMP 4S language versions were examined: Spanish, French, German, Chinese Simplified, Japanese, Russian, and Italian.

The Avant STAMP 4S test is a validated, standardized, computer-adaptive test that is aligned with the ACTFL proficiency guidelines and that assesses a test-taker's language proficiency across the four domains of Reading, Writing, Listening, and Speaking. It is an officially approved test for awarding the State and Global Seals of Biliteracy in all states that offer the seal and is available in 14 languages at the time of this writing.

For the Reading and Listening sections of the Avant STAMP 4S, test-takers answer a suite of adaptive test questions about various passages in the target language and receive a STAMP score between 1 (Novice-Mid) and 9 (Advanced-High), which is aligned to the ACTFL proficiency scale. For the Writing and Speaking sections, test-takers must provide a response to three separate prompts, custom selected depending on the test-taker's results in the Reading and Listening section, respectively. Each of their responses is scored by experienced human raters trained on the ACTFL proficiency scale and guidelines and receives a STAMP level between 1 (Novice-low) and 8 (Advanced-Mid). The combination of the STAMP levels received on the three prompts in the section determine the test-taker's final STAMP proficiency level for the section.

As noted previously, the focus of this paper is on languages that do not benefit from having a validated, four-skill proficiency test such as the Avant STAMP 4S available. These include LCTLs such as Filipino (Tagalog), Czech, Hmong, Vietnamese, Hawaiian, Haitian-Creole, and several others currently available as a two-skill productive proficiency test through the Avant STAMP WS test. Since the Avant STAMP 4S test and the Avant STAMP WS test are built and rated to the same ACTFL standards and developed by the same language testing company, the results observed in the Avant STAMP 4S test are naturally expected to hold for test-takers of the Avant STAMP WS as well.

### 3.3 Participants

The study participants were randomly selected from test-takers who took all four sections of the Avant STAMP 4S test in each of the seven languages between April 2019 - April 2021 and scored at least a STAMP level 5 (Intermediate-Mid) on the Writing and Speaking sections. The reason for using Intermediate-Mid as the minimum score in Writing and Speaking for the

analysis is because that is the proficiency level most often required by states when awarding the State Seal of Biliteracy. It is also the minimum proficiency level required for being awarded a Global Seal of Biliteracy (Intermediate-Mid = Functional Proficiency). The Global Seal of Biliteracy also awards a seal at the Advanced-Low (STAMP 7) level, deemed *Working Proficiency* and at the Advanced-High level for *Professional Fluency*. Therefore, considering the scores only of test-takers who achieved at least a STAMP 5 (Intermediate-Mid) in both Writing and Speaking increases the power of the analysis by focusing on the actual target population of the State and Global Seals of Biliteracy.

The number of test-takers randomly selected for the purposes of this study are:

*Spanish* -> 13,725 test-takers

*French* -> 1,498 test-takers

*German* -> 410 test-takers

*Chinese Simplified* -> 811 test-takers

*Japanese* -> 889 test-takers

*Russian* -> 1,250 test-takers

*Italian* -> 755 test-takers

The results in the dataset were kept intact in order to preserve their integrity. Therefore, even extreme and very rarely observed cases in which a test-taker's receptive score was several STAMP levels below their productive score were kept for the analysis.

### **3.4 Results**

***Research Question 1: What is the correlation between Writing - Reading scores and between Listening - Speaking scores on the Avant STAMP 4S test, across various languages?***

If the correlation between the Writing and Reading scores on the Avant STAMP 4S, as well as between the Speaking and Listening scores is determined to be weak, it does not make much sense to conduct a more fine-grained analysis that would answer research questions 2 and 3. The correlations observed in the dataset for this study, across seven representative languages, can be found in Figure 6.

<b>STAMP 4S Language</b>	<b>Correlation between Reading / Writing scores</b>
<b>Spanish</b>	<b>0.65</b>
<b>French</b>	<b>0.68</b>
<b>German</b>	<b>0.66</b>
<b>Chinese Simplified</b>	<b>0.70</b>
<b>Japanese</b>	<b>0.75</b>
<b>Russian</b>	<b>0.57</b>
<b>Italian</b>	<b>0.70</b>

Figure 6. *Correlation between ACTFL-aligned Reading and Writing scores on the Avant STAMP 4S across seven different languages.*

As defined by Cohen (1988), a strong correlation is defined as one above 0.5. As can be seen in Figure 6 above, we observe a strong correlation between Writing and Reading scores on STAMP 4S for all seven languages in the Avant STAMP 4S sample.

Figure 7 displays the correlation between the Avant STAMP 4S Speaking and Listening scores in our dataset (same test-takers, since only complete data was used in the study):

STAMP 4S Language	Correlation between Listening / Speaking scores
Spanish	0.69
French	0.79
German	0.78
Chinese Simplified	0.71
Japanese	0.83
Russian	0.69
Italian	0.73

Figure 7. Correlation between ACTFL-aligned Listening and Speaking scores on the Avant STAMP 4S across seven different languages.

Just as was observed with the Writing/Reading scores, the correlation between the Speaking and Listening scores on the Avant STAMP 4S is strong for all seven languages in the dataset.

Now that it has been determined that a strong enough correlation exists on the Avant STAMP 4S between the productive skills of Writing and Speaking, and their receptive counterparts, namely Reading and Listening, we turn to examining the extent to which productive scores can be used as a proxy for minimum receptive scores for the purposes of awarding a State or Global Seal of Biliteracy.

**Research Question 2: Could a test-taker’s Writing score be reliably used as indirect evidence (i.e., proxy) of the minimum Reading score the same test-taker would achieve for the purposes of the State or Global Seals of Biliteracy across various languages?**

In order to answer Research Question 2, the Writing score (STAMP level) for each test-taker in the dataset was subtracted from his or her Reading score. Therefore, the result of this subtraction (see on the x/horizontal axis in the figures that follow) is:



- Negative (- 4, - 3, - 2, - 1) in cases where the Reading score was *lower* than the Writing score, with the resulting number indicating the size of the gap between the scores in the two sections.
- Neutral (0) in cases where the Writing and Reading scores were at the *same level* of proficiency
- Positive (+ 1, + 2, + 3, + 4) in cases where the Reading score for a given test-taker was *higher* than their Writing score.

In Figures 8 through 14, we can see the results for all seven languages examined:

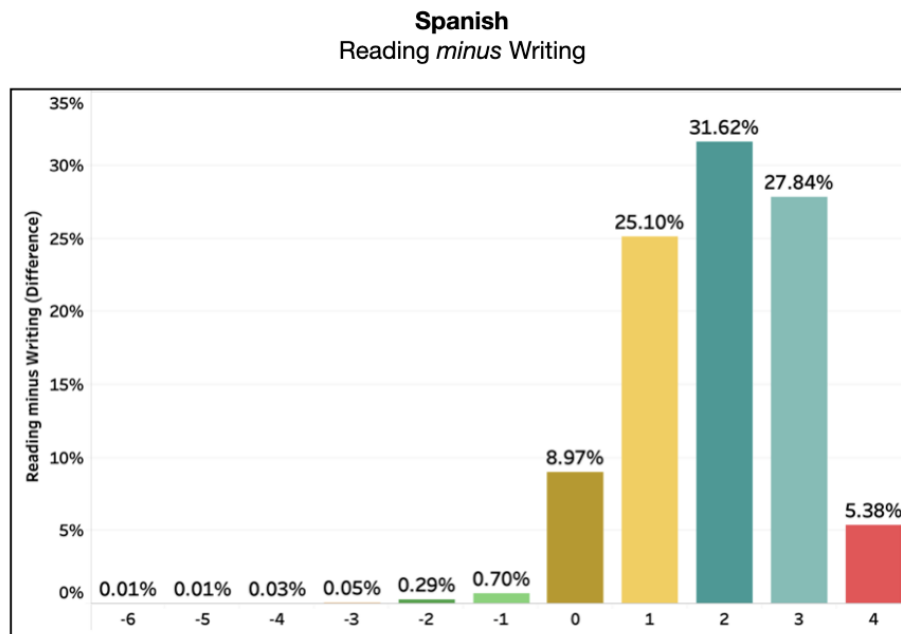


Figure 8. Test-takers' Reading Score *minus* their Writing Score for the Avant STAMP 4S Spanish.

**French**  
Reading *minus* Writing

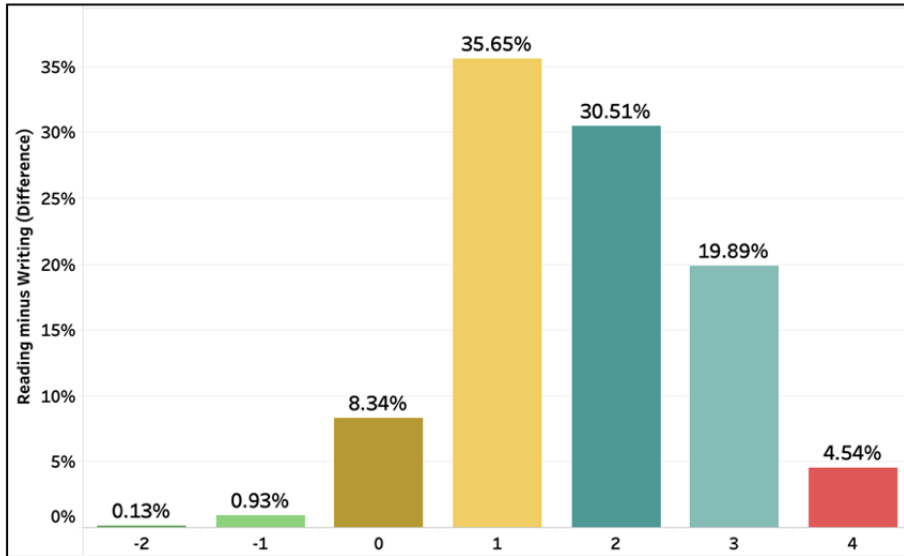


Figure 9. Test-takers' Reading Score minus their Writing Score for the Avant STAMP 4S French.

**German**  
Reading *minus* Writing

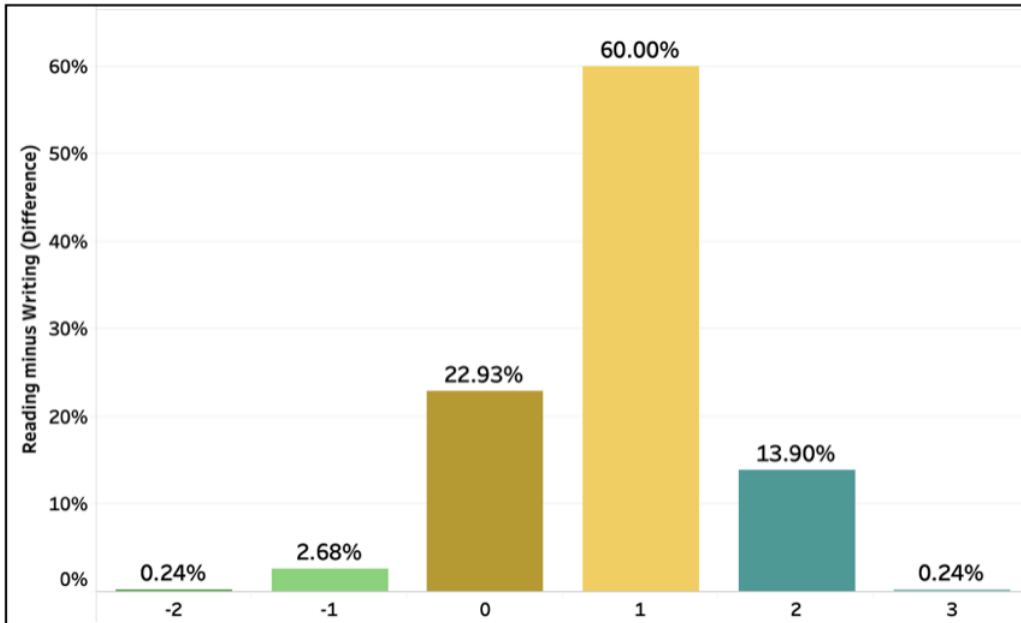


Figure 10. Test-takers' Reading Score minus their Writing Score for the Avant STAMP 4S German.

**Chinese Simplified**  
Reading *minus* Writing

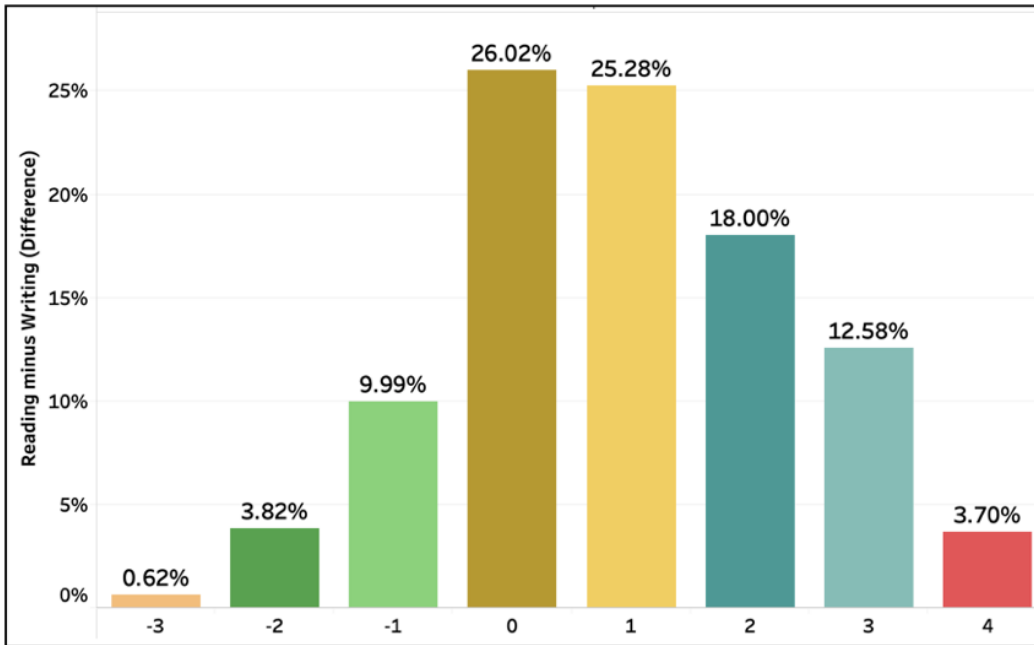


Figure 11. Test-takers' Reading Score *minus* their Writing Score for the Avant STAMP 4S Chinese Simplified.

**Japanese**  
Reading *minus* Writing

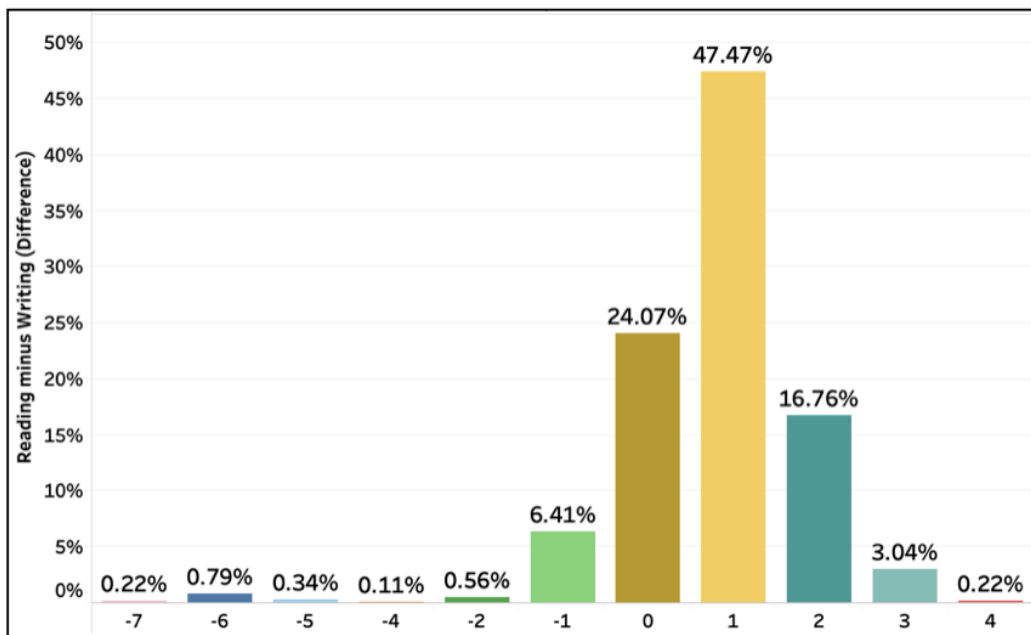


Figure 12. Test-takers' Reading Score *minus* their Writing Score for the Avant STAMP 4S Japanese.

**Russian**  
Reading *minus* Writing

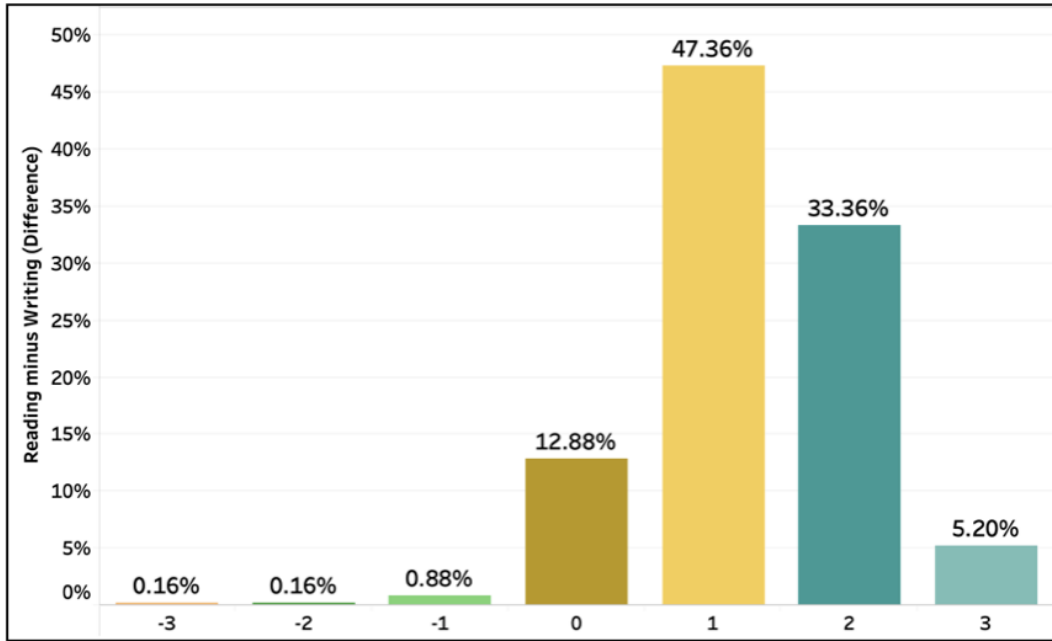


Figure 13. Test-takers' Reading Score minus their Writing Score for the Avant STAMP 4S Russian.

**Italian**  
Reading *minus* Writing

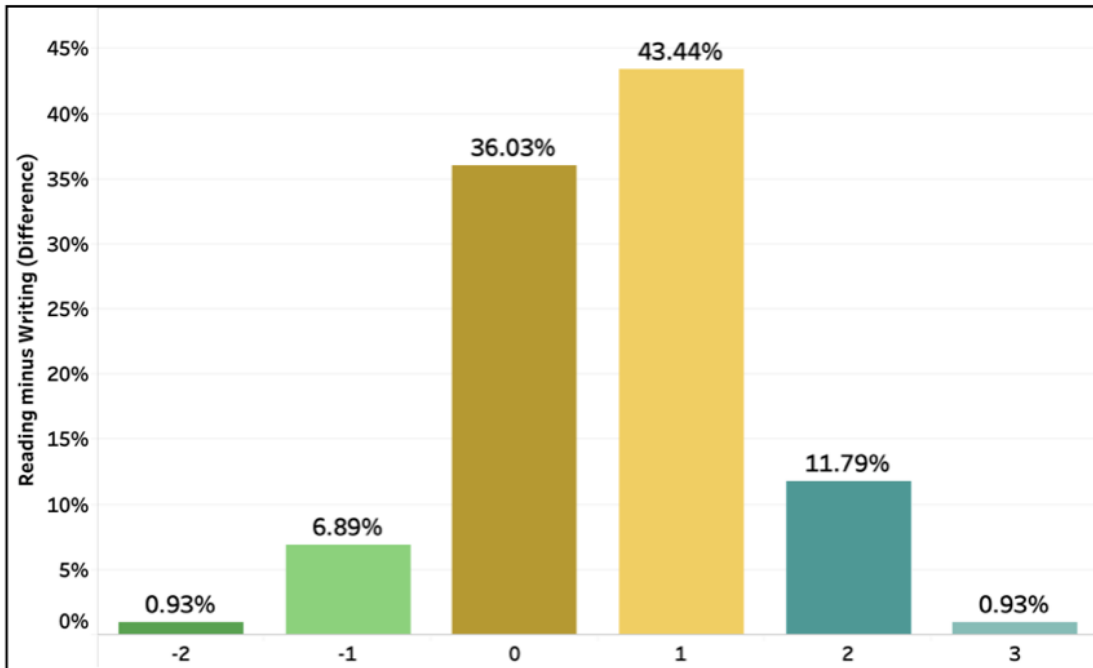


Figure 14. Test-takers' Reading Score minus their Writing Score for the Avant STAMP 4S Italian.

In Figure 15 we indicate, for each of the seven languages, the percentage of test-takers whose Reading score was at a level of proficiency at least as high as that observed for their Writing score:

STAMP 4S language	Percentage of Cases where Reading score $\geq$ Writing score
Spanish	98.91%
French	98.93%
German	97.07%
Chinese Simplified	85.58%
Japanese	91.56%
Russian	98.80%
Italian	92.19%

Figure 15. *Percentage of test-takers with Reading scores at least as high as their Writing scores.*

As can be seen in Figure 15 above, the percentage of cases in which test-takers' Reading score was at least as high as their Writing score varied from 85.58% (Chinese Simplified) to 98.93% (French), with an average, across the seven languages, of 94.72% of the test-takers in the analysis having a Reading score that was at least as high at their Writing score.

Even for a language such as Chinese Simplified, with a more complex, logographic writing system, the great majority of test-takers achieved a Reading score that was in fact higher than their Writing score.

The results above provide strong support to the hypothesis that Writing Scores can be reliably used as indirect evidence of the *minimum* Reading score that test-takers of either the Avant STAMP 4S or the Avant STAMP WS would achieve for the purposes of the State or Global Seals of Biliteracy across several languages.

**Research Question 3: Could a test-taker's Speaking score be reliably used as indirect evidence (i.e., proxy) of the minimum Listening score the same test-taker would achieve for the purposes of the State or Global Seals of Biliteracy across various languages?**

In order to answer Research Question 3, the Speaking score (STAMP level) for each test-taker in the dataset was subtracted from his or her Listening score. The interpretation of the result of this subtraction (see on the x/horizontal axis in Figures 16 through 22) is the same as for Research Question 2.

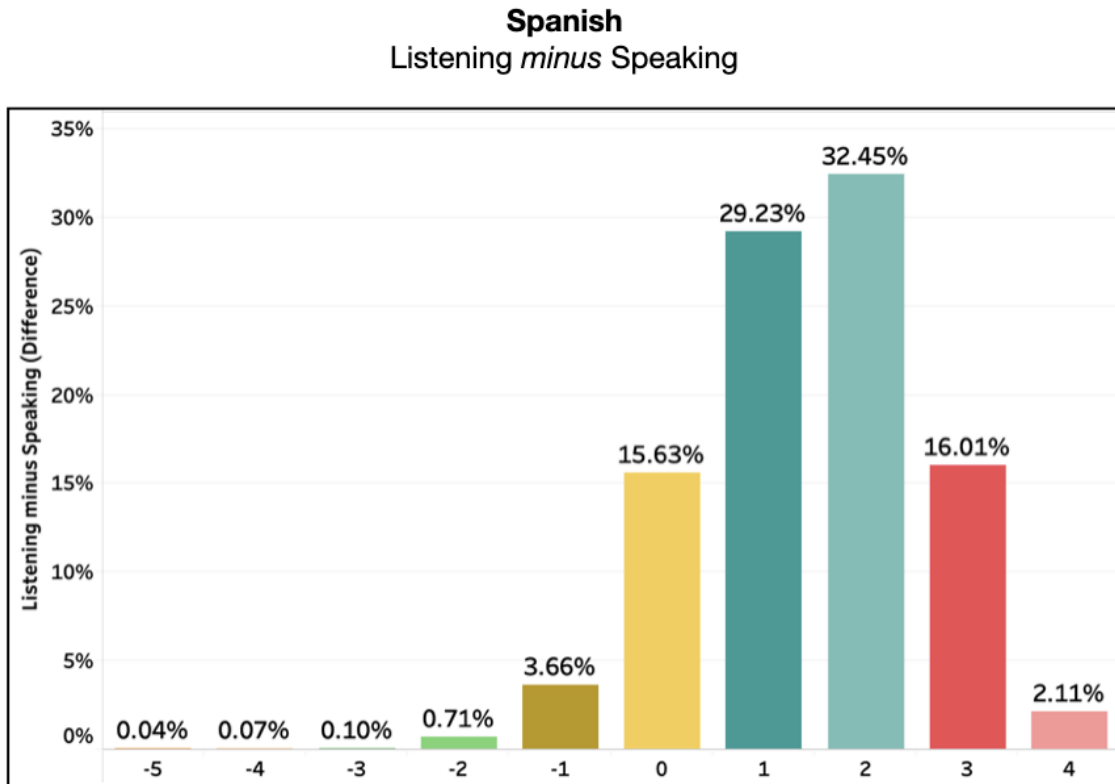


Figure 16. Test-takers' Listening score minus their Speaking score for the Avant STAMP 4S Spanish.

**French**  
Listening *minus* Speaking

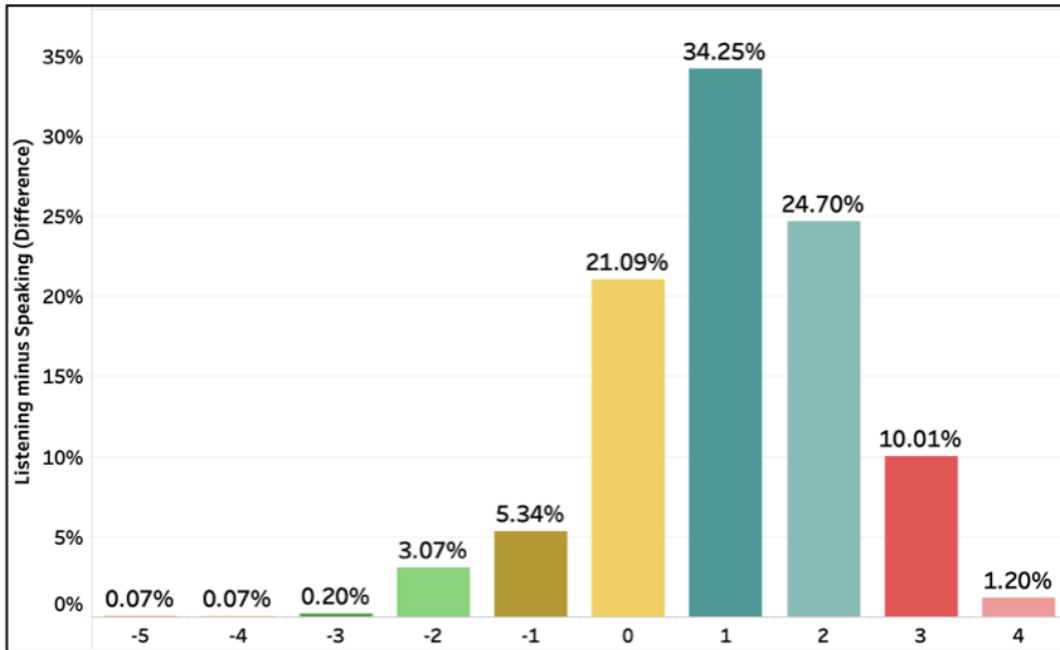


Figure 17. Test-takers' Listening score minus their Speaking score for the Avant STAMP 4S French.

**German**  
Listening *minus* Speaking

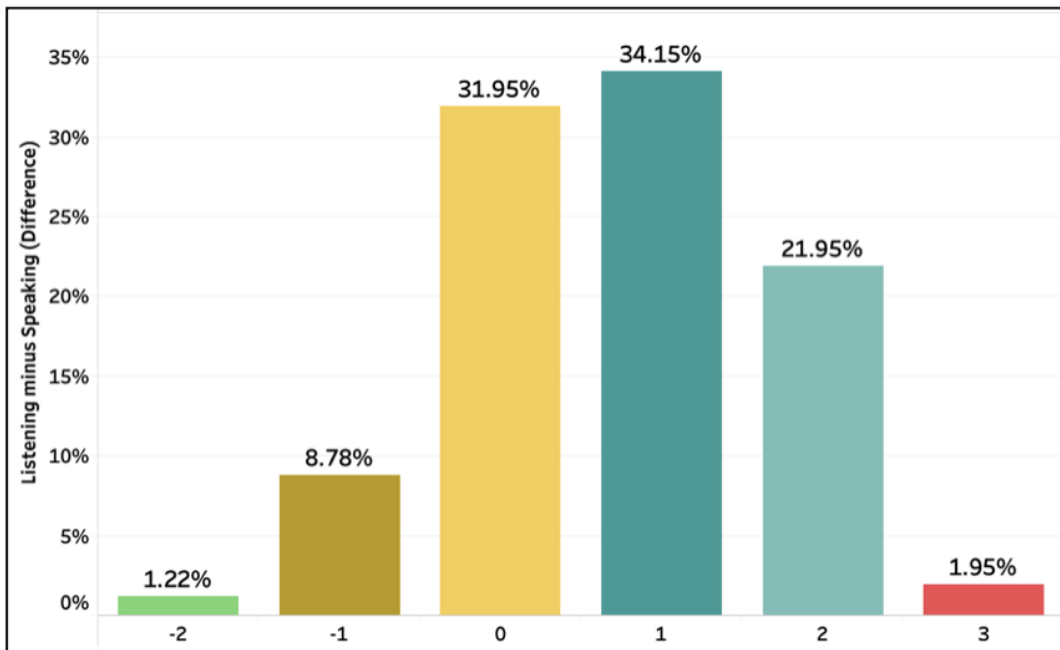


Figure 18. Test-takers' Listening score minus their Speaking score for the Avant STAMP 4S German.

**Chinese Simplified**  
Listening *minus* Speaking

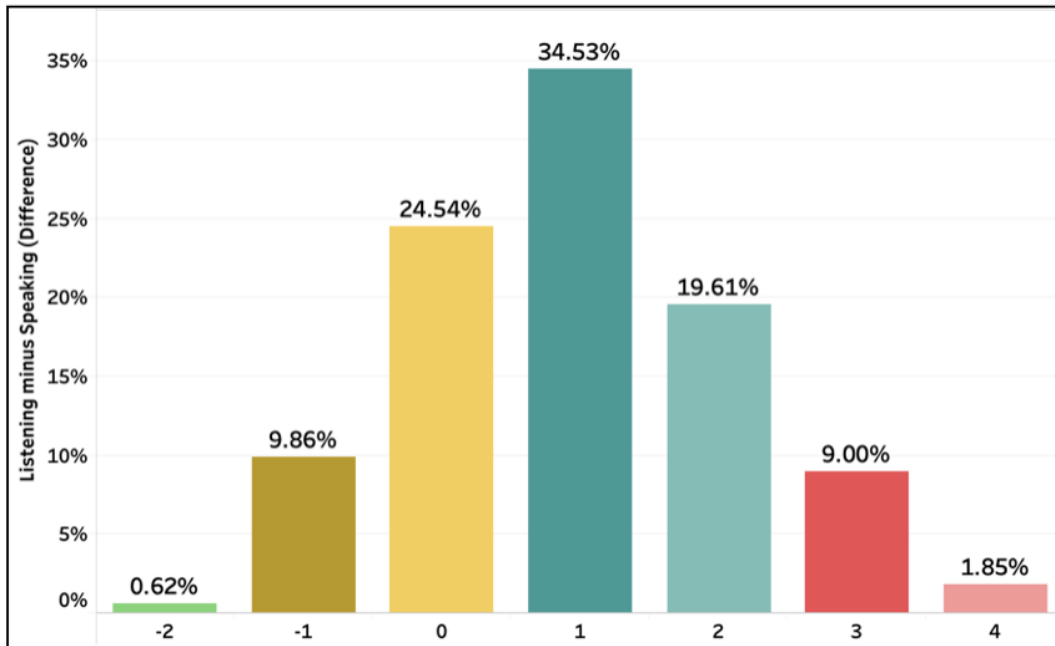


Figure 19. Test-takers' Listening score minus their Speaking score for the Avant STAMP 4S Chinese Simplified.

**Japanese**  
Listening *minus* Speaking

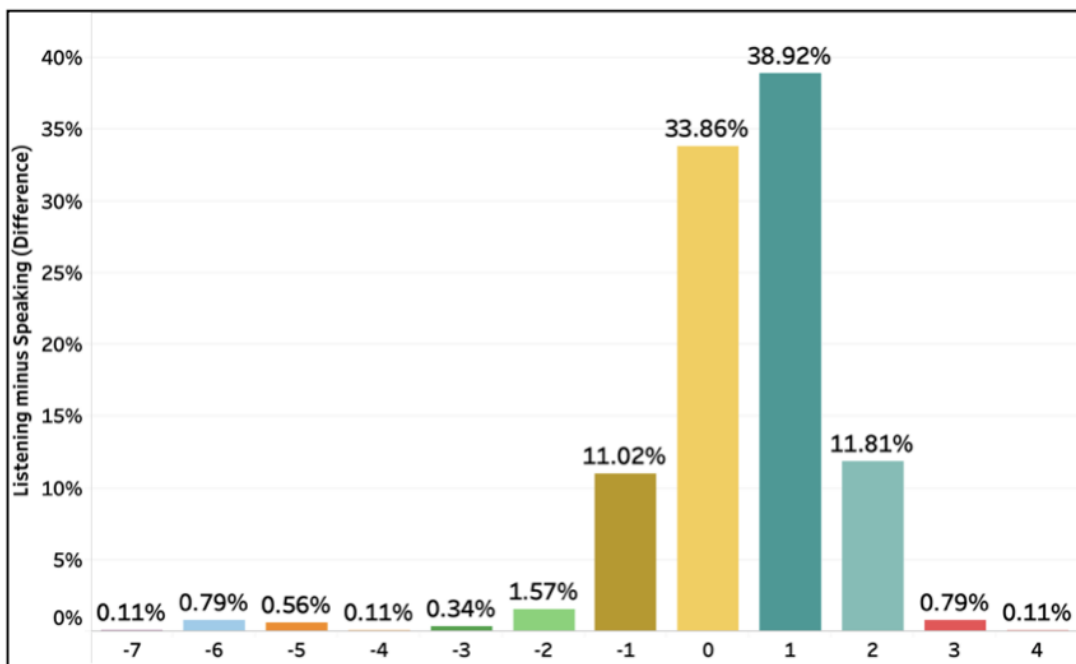


Figure 20. Test-takers' Listening score minus their Speaking score for the Avant STAMP 4S Japanese.



**Russian**  
Listening *minus* Speaking

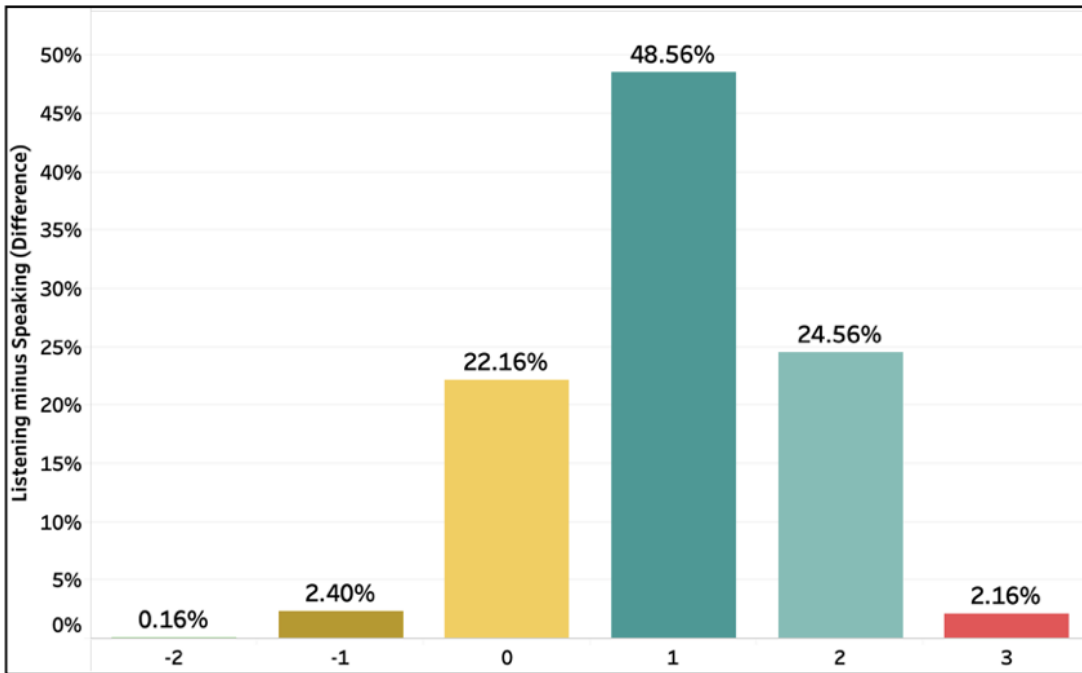


Figure 21. Test-takers' Listening score minus their Speaking score for the Avant STAMP 4S Russian.

**Italian**  
Listening *minus* Speaking

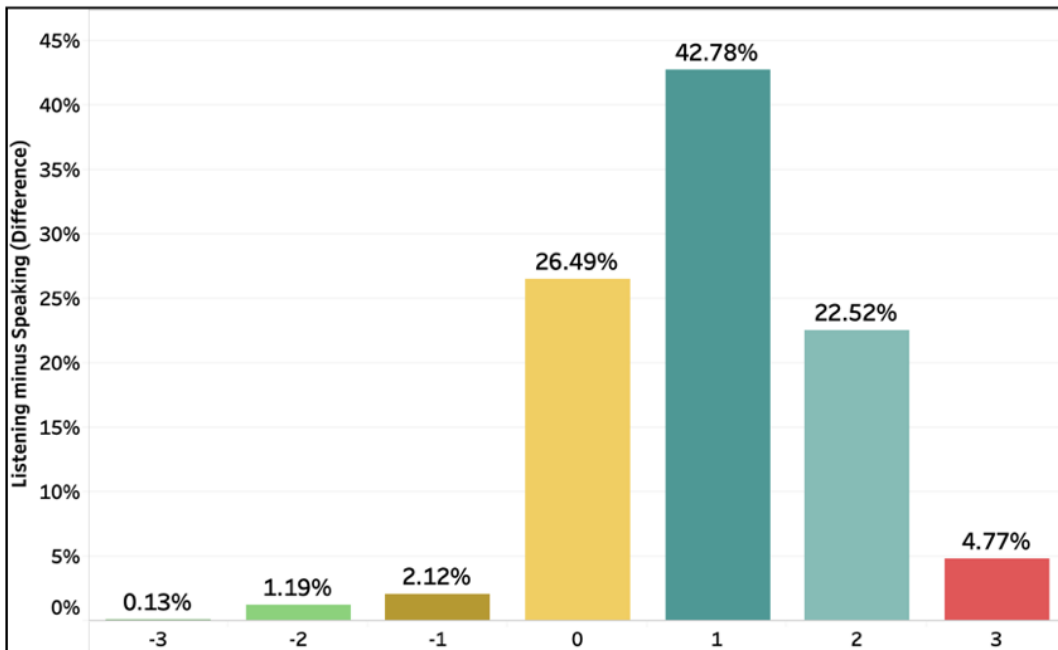


Figure 22. Test-takers' Listening score minus their Speaking score for the Avant STAMP 4S Italian.

In Figure 23 we indicate, for each of the seven languages, the percentage of test-takers whose Listening score was at a level of proficiency at least as high as that observed for their Speaking score:

STAMP 4S language	Percentage of Cases where Listening score $\geq$ Speaking score
Spanish	95.43%
French	91.25%
German	90.00%
Chinese Simplified	89.52%
Japanese	85.49%
Russian	97.44%
Italian	96.56%

Figure 23. Percentage of test-takers with Listening scores at least as high as their Speaking scores.

As can be seen in Figure 23 above, the percentage of cases in which test-takers' Listening score was at least as high as their Speaking score varied from 85.49% (Japanese) to 97.44% (Russian), with an average, across the seven languages, of 92.24% of the test-takers in the analysis having a Listening score that was at least as high at their Speaking score.

The results above provide strong support to the hypothesis that Speaking Scores can be reliably used as indirect evidence of the minimum Listening score that test-takers of either the Avant STAMP 4S or the Avant STAMP WS would achieve for the purposes of the State of Global Seals of Biliteracy across several languages.

## 4. Discussion

The existence of the State Seal of Biliteracy (SoBL) in 42 US States and the District of Columbia at the time of this writing is something to be widely celebrated. It serves as a major step towards recognizing the importance of *all* languages and their speakers, the importance of bilingualism and biliteracy for the individual and for society as a whole, and it helps to acknowledge that biliteracy and fluency in another language is an asset to be celebrated. As is frequently the case with many well-intended policies, however, the SoBL still has several challenges that need to be overcome in order to fulfill its true potential.

One of the current challenges associated with the SoBL is that it is implemented with different standards in different states. Whereas some states only accept a small number of standardized four-skill assessments as a means to demonstrate proficiency and biliteracy in the language, others accept alternative methods of “meeting the qualifications”, including seat time, portfolios, and other substantially less reliable means of assessing a student’s proficiency in the language. Other states, such as Massachusetts, accept a standardized and validated two-skill assessment or Writing and Speaking such as the Avant STAMP WS in order to award the SoBL to as many qualified students as possible, thus increasing the linguistic and demographic diversity of SoBL recipients, and decreasing the costs these states and their districts would incur if they had to produce their own assessments for less commonly taught languages (LCTLs) that do not benefit from the existence of a standardized and commercially available four-skill test at the moment. Even if states and districts could afford to produce their own assessments, there is no guarantee that the end result would be a reliable, fair, unbiased, and accessible test. Maintaining the high quality and validity of an assessment is a major effort and one towards which even large testing organizations must dedicate a substantial amount of human and financial resources on an ongoing basis.

Just as states such as Massachusetts have increased accessibility to the state SoBL through their acceptance of rigorous and reliable two-skill assessments when a four-skill assessment is not available, the Global Seal of Biliteracy has accomplished a very similar goal since its inception in 2018 by accepting validated two- or three- skill assessments in languages such as Bulgarian, Gujarati, Norwegian, Filipino, Swahili, Tamil, Ukrainian, and many others when a four-skill assessment is not available. By awarding its seal to students not only from public schools, but also in private schools, colleges, and students being homeschooled, the Global Seal of Biliteracy has been a major force in bringing increased equity and access to speakers of as many languages as possible, without compromising the rigor that qualifying assessments must show. As evidenced in Figure 5 above, the acceptance of validated, reliable, and commercially available two-skill assessments of proficiency and biliteracy in a language would allow many US states to quickly increase access to the SoBL and therefore make the seal more equitable, provided that the data from these potential two-skill tests support the decisions to be made based on their scores.

The results of the Avant Assessment study of 19,338 test-takers across seven languages, including character-based languages, offers a powerful rationale for accepting a 2-skill test such as the Avant STAMP WS for the purposes of the State or Global Seal of Biliteracy when a standardized and validated 4-skill test is unavailable or impractical. An average of 94.72% of test-takers in the SoBL range had a Reading score at or above their Writing score across the seven languages and an average of 92.24% had a Listening score at or above their Speaking score in the study. By accepting a two-skill assessment of productive skills such as the Avant STAMP WS for the purposes of the State or Global Seal of Biliteracy, we are not trying to identify a test-taker’s exact level of proficiency in Reading and Listening, but simply assessing whether their Writing and Speaking scores allow us, with a very high level of certainty, to make a judgement as to whether their Reading and Listening skills would meet the *minimum* requirements for being awarded the seal.

We believe that the strong results above more than support the policy of both the Global Seal of Biliteracy and the states that accept a 2-skill test such as the Avant STAMP WS for a SoBL, while also noting that the results observed above are a direct result of the specific rigor and procedures employed by Avant Assessment in the development of its tests. While the data from other two-skill assessments may also support the assumption that productive scores in Writing and Speaking are a reliable indicator of a student’s minimum Reading and Listening skills, it is up to different test providers to show that their own test results and data would support the use of their two-skill assessments for the purposes of awarding a State or Global Seal of Biliteracy, just as Avant Assessment has done in this study.

It is our hope that the increased acceptance of well-supported two-skill assessments of language proficiency for the purposes of the State and Global Seals of Biliteracy will significantly increase the equity and fairness of these great initiatives, while also motivating a higher number of language testing organizations to develop assessments in less commonly taught/tested languages that do not yet benefit from a two-skill assessment such as the Avant STAMP WS. This will hopefully bring us closer to ensuring that the goal of the seal, namely to “recognize a student who has attained proficiency in English and one or more other world languages by high school graduation” applies to as many students as possible, including as many English Learners (ELs) as possible, giving them an additional reason to sustain and grow their linguistic and cultural assets.

#### REFERENCES:

[ACTFL, NABE, NCSSFL & TESOL \(2015\). Guidelines for Implementing the Seal of Biliteracy.](#)

[ACTFL \(2012\). ACTFL Proficiency Guidelines 2012.](#)

[Berninger, V. W. \(2000\). Development of Language by hand and its connections with language by ear, mouth, and eye. Topics in Language Disorders, 20\(4\), 65–84.](#)

[Black, Chou & Hancock \(2020\). The 2018-19 National Seal of Biliteracy Report.   
www.SealofBiliteracy.org](#)

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

[David, K. J., & Heineke, A. J. \(2017\). The Seal of Biliteracy: Variations in policy and outcomes. Foreign Language Annals, 50\(3\), 486–499.](#)

Greenberg Motamedi, J., & Jaffery, Z. (2015). The impact of competency-based world language credits on graduation: A case study from the Washington Road Map districts. Portland, OR: Education Northwest.

[Heineke, A. J., Davin, K. J., & Bedford, A. \(2018\). The Seal of Biliteracy: Considering Equity and Access for English Learners. \*education policy analysis archives\*, 26\(99\), n99.](#)

Hulstijn, J. H. (2015). *Language proficiency in native and non-native speakers. Theory and research* (Vol. 41). John Benjamins Publishing Company.

[Illinois General Assembly \(2013\). Senate Bill 1221.](#)

[Kiersz, A., De Luce, I., & Hoff, M. \(2020\). This Map Shows the most commonly spoken language in every US state, excluding English and Spanish. Business Insider, Aug 12, 2020.](#)

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: do we need both to measure vocabulary knowledge? *Language Testing*, 21(2), 202–226.

[National Center for Education Statistics \(2021\). English Language Learners in Public Schools.](#)

[Pearson \(2018\). Test Comparisons: A Comparison of Three Internationally Recognized Tests of Spoken English.](#)

Powers, Donald E. (ed.) (2013) The Research Foundation for the TOEIC Tests: A Compendium of Studies: Volume II. Princeton, NJ: Educational Testing Service, Sep 2013, p1.1-1.7

Schoonen, R. (2019). Are reading and writing building on the same skills? The relationship between reading and writing in L1 and EFL. *Reading and Writing*, 32(3), 511-535.

Richards, J. (2015). Bridging the gap between receptive and productive competence. World of Learning blog. Cambridge University Press.

[U.S. Department of Education \(2018\). Our Nation's English Learners.](#)

Wagner, E. (2016). A study of the use of the TOEFL iBT® test speaking and listening scores for international teaching assistant screening. ETS Research Report Series, 2016(1), 1-48.