

The Reliability and Validity of Language Proficiency Assessments for English Language Learners

Xiaoyan Jing

Shandong Shengli Vocational College, Dongying, Shandong, 257097, China

Abstract: English language learners (ELLs) have become one of the fastest growing groups in elementary schools in the United States of America. This paper will research the English proficiency assessments currently used for ELLs in the early-exit TBE programs in elementary schools and the validity and reliability of these English language proficiency assessments.

Keywords: reliability; validity; ELLs

DOI: 10.36012/fhe.v1i1.893

1. Introduction

English language learners (ELLs) have become one of the fastest growing groups in elementary schools in the United States of America. According to the National Center for Education Statistics in 2010, English language learners (ELLs) are increasing in the public schools. It is estimated that in 2011 to 2012, 4.4 million students in the U.S., which accounted for 9.1% of the school-aged population, were classified as ELLs while the number of ELLs varies in states, ranging from 0.7% (West Virginia) to 23.2% (California) (Carroll & Bailey, 2016). The Hispanic ELLs account for 90% of all the ELLs in the United States (Umansky & Reardon, 2014). Albers, Kenyon, and Boal (2009) stated that ELLs academically lag behind the non-ELLs, and have “higher rates of negative occupational, economic, social, and physical outcomes” (p. 76).

According to the mandatory federal requirements of No Child Left Behind (NCLB) in 2001, speaking, writing, listening, and reading proficiencies of all English language learners in K-12 should be assessed. Kenyon, MacGregor, Li and Cook (2011) stated that “One of the mandates of the No Child Left Behind Act is that states show adequate yearly progress in their English language learners’ (ELLs) acquisition of English language proficiency. Traditionally language ability has been composed of four skills, listening, reading, speaking, and writing (Bachman & Palmer, 2010). States are required to assess ELLs’ English language

proficiency annually in these four language domains (listening, reading, writing, and speaking) to measure their progress; they are also required to report on “a composite comprehension measure” (Kenyon et al., 2011, p. 383). The test results will be used to establish “Annual Measurable Achievement Outcomes (AMAOs), which are reported by states to the federal government” (Kenyon et al., 2011, p. 384). No Child Left Behind also required that it is mandatory for English language learners to participate in large-scale assessments.

There are researches focused on the large-scale assessments with the main stream students. However, it is worth examining these proficiency assessments the ELLs take and the validity, reliability and effectiveness of these assessments as they play a crucial role in measuring the English proficiency of ELLs and classifying the ELLs into different levels and deciding whether the ELLs can exit the programs or not.

Gottlieb and Nguyen (2007) listed five purposes to assess the language proficiency of students in different language education programs:

Contribute to the identification and placement of ELLs. Monitor the progress of ELLs’ English language development to inform teaching and differentiate instruction. Document growth of English language development over time for state accountability. Contribute to the reclassification of ELLs when deemed English language proficient. Contribute to the evaluation of language program services (Gottlieb & Nguyen, 2007, p. 33).

2. Significance

Bachman and Palmer (2010) stated that the consequences of language assessment will influence not only the individuals but also the educational system and society. Former researches have proved that the ELLs have been misplaced and improperly classified in the English education programs or even in the special education. Solano-Flores (2008) believed that the current tests used for ELLs are based on "deterministic views" and "erroneous assumptions" (p. 189). For example, the ELLs are classified into a few language proficiency categories, and only the English proficiency of the ELLs is assessed without any assessment of their first language proficiency.

Most of the bilingual programs focus mainly on English proficiency development while failing to show their first language proficiency development. Second, Solano-Flores (2008) also stated that the assessments for ELLs are rarely comprehensive enough to show their listening, speaking, reading and writing development of first language and second languages. "As a result, language tests may provide fragmented and inconsistent information about the linguistic proficiency of ELLs" (p. 189). Third, the definition of ELLs is related to their demographic characteristics and tends to misclassify the ELLs. The English language proficiency tests for ELLs are designed according to the monolingual, English-speaking students instead of the ELLs who learn English as a second language, which also leads to misclassification of the ELLs.

To examine the English proficiency assessments can, in the short term, reduce the misplacement of ELLs, improve the English proficiency assessment of ELLs, help ELLs exit the program and help to make decision and policy. In the long run, improvement of the English proficiency tests for ELLs can guide bilingual teachers to interpret English proficiency test scores and assess the ELLs effectively, compare the English proficiency tests currently used and explore the effectiveness of these English proficiency tests, and to search the more effective methods to assess the ELLs' English proficiency.

The following research questions will be addressed in this library-based research paper:

1. What are the English proficiency assessments currently used for ELLs in the early-exit TBE programs in elementary schools?

2. What are the validity and reliability of these English language proficiency assessments?

3. Review of the Literature

There are studies explained the different English language proficiency tests adopted to test the ELLs (Albers et al., 2009), which compared the pro-NCLB tests and post-NCLB tests. There are also studies defined the construct in the English language proficiency tests (Bailey & Huang, 2011), and the influences exerted by the score report after the English language proficiency tests (Abedi, 2008). The study conducted by MacSwan and Pray (2005) focused on the linguistic requirements in the assessments while Bailey (2017) proposed integration of linguistic requirements with academic content requirements.

MacSwan and Pray (2005) examined 89 ELLs in bilingual programs and found out that they did not learn English slower than students in only English programs and they believed that English proficiency should involve phonology, morphology, semantics, syntax and pragmatics. Bailey (2007) integrated the linguistic requirements with the academic English skills to assess the ELLs' linguistic ability and academic content, thus forming the academic English proficiency assessment, including, listening, reading, speaking and writing.

Hauck, Wolf and Mislevy (2016) stated that NCLB changed the education of ELLs from two aspects. First, the ELLs are required to be included in the standardized assessments as the non-ELLs to meet the accountability requirement of NCLB. Second, it is required that new English language proficiency standards should be established by every state which is supposed to be corresponding to content standards, and the standardized English language proficiency assessments should test the ELLs annually.

3.1 The Current English Proficiency Assessments Used in the Bilingual Program

Before the No Child Left Behind in 2002, there were also English language proficiency assessments in different states. However, they were not uniform and mainly focused on the social language proficiency of ELLs (Albers et al., 2009).

Albers et al. (2009) compared the pre-NCLB ELP tests in which social language proficiency was tested, while post-NCLB ELP tests in which academic language proficiency was included in the tests. Whether

ELP tests can benefit classroom instructions received less attention. The fundamental ELP standards include communication for social and instructional purposes, and the communication of "information, ideas and concepts for concept success" (Albers et al., 2009, p. 77) in content areas of language arts, mathematics, science and social studies.

Findings showed that ACCESS for ELLs, though measuring the similar constructs as the pre-NCLB ELP tests, has components not cover in these pre-NCLB ELP tests. ACCESS for ELLs, as a new-generation ELP measurement, incorporated the academic language proficiency to utilize the data obtained, enhance the instructions in the classrooms, and provide interventions to the ELLs (Albers et al., 2009).

According to Wolf, Kao, Griffin, Herman, Bachman, Chang, and Farnsworth (2008), English language learners are linguistically and culturally varied, and with diverse academic levels and language proficiency, while 44% of the ELLs are in pre-kindergarten to Grade 3. According to the NCLB, "a home language survey, an ELP assessment, and academic achievement assessment(s) in content areas" are employed to identify ELL students (Wolf, et al., 2008, p. 3).

Wolf et al., (2008) explained the consortia under the NCLB and stated that "nearly 40 states were originally represented in the four consortia, which are known as Mountain West Assessment Consortium (MWAC), Pennsylvania Enhanced Assessment Grant (PA EAG), State Collaborative on Assessment and Student Standards (SCASS) Consortium, and World-Class Instructional Design and Assessment (WIDA) Consortium" (p. 6). The English language proficiency assessments have developed based on these consortia. There are some states that do not use these assessments cooperated with test publishers to assess the ELLs. For example, WIDA standards are the basis and theoretical constructs for the "Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs)" (p. 13) with the purpose to determine the language proficiency of ELLs and place them into different programs with different language support.

Fox and Fairbairn (2011) examined Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs), which is "a large-scale, high-stakes,

standards-based, and criterion-referenced English language proficiency test administered in the USA annually to more than 840,000 English Language Learners (ELLs), in K-12 classrooms" (p. 425). ACCESS for ELLs is a test system developed by World-Class Instructional Design and Assessment (WIDA) Consortium by three states, Wisconsin, Delaware, and Arkansas to meet the assessment requirements in the NCLB Act. ACCESS for ELLs first, used a vertical scaling to show the ELLs' proficiency in listening, speaking, reading and writing in the different grades starting from the first grade. The vertical scaling enables to demonstrate even the small differences in the ELLs' performances, especially when they move across different grade levels (Fox & Fairbairn, 2011).

According to Bunch (2011), the four states, Pennsylvania, Utah, Nevada, and Wisconsin were funded under NCLB and developed the English proficiency tests to meet the requirement of NCLB. They all have reading, listening, reading and writing tests, and employ various types of test items. There is an oral test in these assessments scored on the basis of standardized rubric. Classical response theory and item response theory are utilized, and formal standardized setting is included. They only differ in the specifics.

According to Kenyon, MacGregor and Cook (2011), it is mandatory for ELLs to take English proficiency tests, and the results of English proficiency tests enable the districts and elementary schools to place ELLs, evaluate the effectiveness of programs and help the ELLs make progress in their English acquisition and exit the ELLs. Solano-Flores (2008) analyzed the problems existed in the testing for ELLs and considered testing as "a communication process between assessment systems and ELLs" (p. 189). Bunch (2011) examined the currently used four commercial tests for ELLs, ACCESS for ELLs, CELLA, ELDA, and MWA. He pointed out that the four tests are similar to each other because they all cover listening, speaking, reading and writing. The test items adopted in these four tests are primarily multiple choice questions and open-ended questions.

3.2 Constructs of English Language Proficiency Assessments

According to American Educational Research Association, American Psychological Association and National Council on Measurement in Education (2014), the construct measured in the tests should be included

in the interpretation, which refers to "the concept or characteristic that a test is designed to measure" (p. 11).

Bailey and Huang (2011) defined academic English construct as "the vocabulary, sentence structures, and discourse associated with language used to teach academic content as well as the language used to navigate the school setting more generally" (p. 347). They examined the standards of the English proficiency tests and examined the role the construct plays in the standards of the English language proficiency test domains. It is mandated by NCLB that standards should be applied in the accountability system of these tests. According to NCLB, one state or collaborating states in one consortium can create their own English language proficiency standards.

Abedi (2008) stated that the different standards used in the English language proficiency assessments produced inconsistent classification of ELLs. In addition, the performance levels of listening, reading, speaking and writing are inconsistent, which makes it difficult to interpret the results of students' performance.

According to Abedi (2008), score report and interpretation can be influenced by the number of constructs measured in the tests. The scores from reading, listening, speaking and writing can be combined when they all measure the same construct. However, it would be difficult to combine the scores when they have different constructs to measure. The English language proficiency assessments currently used integrate studies about psychometric aspects and validation of the assessments. For example, "a criterion-related validity approach ... content validation through alignment to ELP content standards and construct validations using the confirmatory factor analytic approach" (Abedi, 2008, p. 200).

Bailey and Huang (2011) emphasized the incorporation of the academic content construct into the English language proficiency tests. First, teachers for non-ELLs can be involved in the teaching of ELLs to help ELLs to take academic content tests. ELL teachers and mainstream class teachers are supposed to cooperate with each other so that linguistic knowledge and content knowledge can be used to improve the standards of the test. According to Bailey and Huang (2011), twenty six states have adopted WIDA standards, published in 2007 and augmented by TESOL in 2006, which provide directions for the ACCESS.

However, there are limited empirical researches about academic English development. Future researches are recommended to validate the current standards, and create new state standards.

Wolf and Faulkner - Bond (2016) proposed that the academic and social language proficiency constructs operated in the English proficiency assessment should be "next-generation" (p. 17) issue. Future researches can examine the alignment between English language proficiency standards and English language proficiency assessments to understand the construct of English language proficiency, which will also influence the social and academic language proficiency included in the English language proficiency.

The constructs measured in the English language proficiency assessment should be explained explicitly to the policy makers to help them choose the proper assessment, and teachers to enable proper instructions to the ELLs. It is also suggested that ELLs from different grades should be included in the future study to examine the relationship between the English language proficiency and academic content assessment, especially in the upper grades (Wolf & Faulkner - Bond, 2016).

3.3 Validity of English Language Proficiency Tests

According to Abedi (2008), to ensure the reliability and validity of the English language proficiency assessments, the academic English is required to be included in the assessments, which also help the ELLs to meet the academic content standards. According to the Department of Education, the link between English language proficiency assessments and academic content assessments is to enable the ELLs to achieve proficiency not only in English but also in other subjects such as math and science. The English language proficiency assessments are on the basis of academic English assessment to facilitate ELLs' content learning.

Wolf and Faulkner - Bond (2016) claimed that English language proficiency assessment is considered to be high-stake and in large scale, the primary and the most important criterion in classifying and exiting ELLs. English language proficiency assessment is not only crucial to ELLs but also to teachers and the school districts due the funding locations distribution. It is vital to examine the validity of the assessment, especially the constructs of the English language proficiency assessment. First, the academic language proficiency has been included as one of the constructs so as

to decide whether the students have been academically prepared to exit the program.

According to American Educational Research Association, American Psychological Association and National Council on Measurement in Education (2014), "validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing tests and evaluating tests" (p. 11). Besides, the construct measured in the tests should be included in the interpretation, which refers to "the concept or characteristic that a test is designed to measure" (p. 11). The following researches will examine the validity of English language proficiency, including "evidence based on test content", "evidence based on response processes", "evidence based on internal structure", and "evidence based on relations to other variables" (p.14-16).

It is vital to test the English language proficiency assessment validity and technical quality because if the assessments about ELLs are inaccurate or improper, the individual ELLs can be inappropriately placed or misclassified. Second, the conclusions about ELLs ELP proficiency will elicit the evaluations about the programs in which ELLs are studying. Subsequently, the programs, schools and even the districts or states will be sanctioned due to these conclusions. Furthermore, the assessments not only measure ELLs' learning but also fundamentally promote reform to improve learning and teaching (Wolf, Farnsworth & Herman, 2008).

Wolf, Guzman - Orth and Hauck (2016) also claimed that it is crucial to investigate the validity of criteria to reclassify and designate the ELLs. Multiple criteria were utilized to reclassify ELLs, which include scores from ELP assessment, content assessment, and local criteria including GPA, grades, and recommendations from teachers. However, the criteria used in reclassification are not suitable or designed to measure language proficiency, which is detrimental to the reclassification of ELLs.

Miley and Farmer (2017) stated that "to ensure the reliability and validity of the WIDA-Access, the assessment is annually field-tested to validate its ability to equitably assess English proficiency" (P. 202). Based on the field testing of Kenyon (2005) in 2004, which was about ACCESS for ELLs, including "five proficiency levels: Entering, Beginning, Developing,

Expanding, and Bridging" (p. 14). There were 6662 participants in this testing, among which 61% were Spanish ELLs. First, sixteen experts and teachers examined whether the content of ACCESS aligned with the construct in WIDA. Evidence based on the ELLs' listening and reading scores showed that test items can assess the English language proficiency of ELLs in these five levels. Then, they examined the interrelations between scores from ACCESS and the four established tests, "IDEA Proficiency Test (IPT), the Language Assessment Scale (LAS), the Language Proficiency Test Series (LPTS), and the Revised Maculaitis II (MAC II)" (p. 15), and found the correlations between them proved the construct validity of ACCESS. "Rasch Modeling" and "Infit and Outfit" (p. 15) statistics were used to examine the four subsets in ACCESS and proved that the internal structure was acceptable. The validity of ACCESS has been approved.

Wolf, Farnsworth and Herman (2008) reviewed 49 states on the systems of their ELP assessment and the according validity. Findings showed that it is essential to align the ELP standards and assessment to ensure the validity. Findings suggested that academic English should be included in the ELP assessments construct. "NCLB legislation stipulates that the constructs of an ELP assessment be aligned with the state's ELP and content standards to measure the progress of appropriate English language development" (Wolf et al., 2008, p. 98).

Wolf and Faulkner - Bond (2016) analyzed the linguistic content of the items in English language assessments in the three states and discovered that social and academic constructs have been represented in the assessment, and in the reading, listening, speaking and writing language domains. Findings showed that the representation of academic language in English language proficiency assessment can support the understanding of ELLs' English language proficiency. Wolf and Faulkner - Bond (2016) adopted the "hierarchical linear modeling" (p. 8) which is to investigate the relationship between the scores from four domains of the language assessment and those from the academic assessment, and revealed that the English language proficiency assessment scores consistently predict the ELLs' performance in content assessment in these three states. The validity of the English language proficiency assessment has been approved. According to Wolf and Faulkner - Bond (2016), this can be ex-

plained by the fact that the two assessments share some English language construct.

Results also demonstrated that social language scores can also be predictive in the performance in content assessment (Wolf & Faulkner - Bond, 2016). ELLs' social language skills are relevant to the progress and proficiency assessment. Except for the constructs taken in the English language proficiency assessment, the various language skills should also be taken into consideration.

3.4 Reliability of English Language Proficiency Assessment

According to Reynolds, Livingston, Willson and Willson (2009), the context of measurement reliability refers to "consistency or stability of assessment results" (p. 91). There are different types of reliability, such as test-retest reliability, alternative-form reliability internal-consistency reliability and inter-rater reliability (Reynolds et al., 2009).

According to Longabach and Peyton (2017), it is of great importance to assess the sub-score reliability and total score reliability when assessing the English proficiency of ELLs because the ultimate goal of assessment is to improve the education for ELLs, which relies on the accuracy of the measurement and appropriate purpose to assessment ELLs. The reliability of English language proficiency tests should also be examined.

Longabach and Peyton (2017) used Cronbach's alpha and standard error of measurement respectively to estimate the reliability and precision of the four methods. Findings showed that CTT and UIRT were similar but lower than A-IRT and MIRT in reliability and precision. MIRT was found to be the most reliable one among the four methods to score the sub-domains for all grade levels. A-IRT followed MIRT closely, ranking second in reliability. These score methods are highly consistent on the sub-scores reported from each grade. The new English language proficiency assessments are being developed in the states, especially the states, such as California, Texas and New York, which claim the large number of ELLs.

Sotelo - Dynega, Ortiz, Flanagan and Chaplin (2013) conducted an empirical research and examined the relationship between the performances of ELLs in the second grade in "New York State English as a Second Language Achievement Test (NYSESLAT)" (p. 787) and the "Woodcock-Johnson III Test of Cognitive

Abilities" to see whether the ELLs' English proficiency is related to their cognitive abilities. NYSESLAT is a language proficiency test developed under the guidance of NCLB for the ELLs, and ELLs' raw scores will be converted to scaled scores. There are four levels "Beginner, Intermediate, Advanced, and Proficient" (p. 787), and it is proved to be reliable and valid. Findings showed that the fact that ELLs are tested through verbal or nonverbal items should be avoided. Instead, the ELLs' current developmental English proficiency should be evaluated because the ELLs' English proficiency has been developed just to cater to the tests, but actually they cannot meet the demands of tests with higher linguistic requirements and cultural backgrounds.

Findings also revealed that the application of verbal or nonverbal tests on ELLs was superficial and the fairness and equity cannot be ensured (Sotelo - Dynega et al., 2013). The ELLs who are tested to have proficient English levels may not be same as the native English speaking peers.

4. Findings

The four commercial tests for ELLs, ACCESS for ELLs, CELLA, ELDA, and MWA were found to be similar to each other including listening, speaking, reading, and writing (Bunch, 2011). ACCESS for ELLs used a vertical scaling to demonstrate the ELLs' proficiency in listening, speaking, reading and writing in the different grades, which enables the demonstration of the small differences in the ELLs' performances, especially when ELLs move across different grade levels (Fox & Fairbairn, 2011).

With the examination of the content, interrelations between scores from ACCESS and four other established tests, and found the correlations between them proved the construct validity of ACCESS. The examination of four subsets, listening, reading, speaking, and writing in ACCESS proved that the internal structure was acceptable. The validity of ACCESS has been approved.

The four scores methods were also compared and examined; these score methods are highly consistent on the sub-scores reported from each grade. According to Longabach and Peyton (2017), MIRT was found to be the most reliable one among the four methods to score the sub-domains for all grade levels.

References

- [1] Abedi, J. Measuring students' level of English proficiency: Educational significance and assessment requirements [J]. *Educational Assessment*,13(2-3): 193-214. doi: 10.1080/10627190802394404
- [2] Albers, C. A., Kenyon, D. M., & Boals, T. J. Measures for determining English language proficiency and the resulting implications for instructional provision and intervention[J]. *Assessment for Effective Intervention*,34(2): 74-85.
- [3] Bachman, L. F., & Palmer, A. *Language assessment in practice*. Oxford University Press.
- [4] Bailey, A. L., & Huang, B. H. Do current English language development/proficiency standards reflect the English needed for success in school [J]. *Language Testing*,28(3): 343-365. doi: 10.1177/0265532211404187
- [5] Bunch, M. B. Testing English language learners under no child left behind[J].*Language Testing*,28(3): 323-341. doi: 10.1177/0265532211404186
- [6] Carroll, P. E., & Bailey, A. L. Do decision rules matter? A descriptive study of English language proficiency assessment classifications for English-language learners and native English speakers in fifth grade [J].*Language Testing*,33(1): 23-52.
- [7] Fox, J., & Fairbairn, S. Test review: ACCESS for ELLs® [J].*Language testing*,28(3):425-431.doi: 10.1177/0265532211404195
- [8] Gottlieb, M. H., & Nguyen, D. *Assessment and accountability in language education programs: A guide for administrators and teachers*. Philadelphia, PA: Caslon.
- [9] Hauck, M. C., Wolf, M. K., & Mislavy, R. Creating a Next - Generation System of K - 12 English Learner Language Proficiency Assessments [J].ETS Research Report Series, 2016(1): 1-10. doi:10.1002/ets2.12092
- [10] Kenyon, D. M., MacGregor, D., Li, D., & Cook, H. G. Issues in vertical scaling of a K-12 English language proficiency test[J]. *Language Testing*,28(3): 383-400. doi: 10.1177/0265532211404190
- [11] Longabach, T., & Peyton, V. (2017). A comparison of reliability and precision of subscore reporting methods for a state English language proficiency assessment[J].*Language Testing*, 1(21). doi: 10.1177/0265532217689949
- [12] MacSwan, J., & Pray, L. Learning English bilingually: Age of onset of exposure and rate of acquisition among English language learners in a bilingual education program [J]. *Bilingual Research Journal*, 29(3): 653 - 678.
- [13] Miley, S. K., & Farmer, A. English Language Proficiency and Content Assessment Performance: A Comparison of English Learners and Native English Speakers Achievement [J]. *English Language Teaching*,10(9): 198. doi: 10.5539/elt.v10n9p198
- [14] Reynolds, C. R., Livingston, R. B., Willson, V. L., & Willson, V. *Measurement and assessment in education*. Upper Saddle River, NJ: Pearson Education International.
- [15] Solano-Flores, G. Who is given tests in what language by whom, when, and where? The need for probabilistic views of language in the testing of English language learners[J]. *Educational Researcher*,37(4): 189-199.
- [16] Sotelo - Dynega, M., Ortiz, S. O., Flanagan, D. P., & Chaplin, W. F. English language proficiency and test performance: An evaluation of bilingual students with the Woodcock - Johnson iii tests of cognitive abilities [J]. *Psychology in the Schools*, 50(8): 781-797. doi: 10.1002/pits.21706
- [17] Umansky, I. M., & Reardon, S. F. Reclassification patterns among Latino English learner students in bilingual, dual immersion, and English immersion classrooms[J].*American Educational Research Journal*,51(5): 879-912. doi:10.3102/0002831214545110
- [18] Wolf, M. K., & Faulkner - Bond, M. Validating English language proficiency assessment uses for English learners: Academic language proficiency and content assessment performance [J].*Educational Measurement: Issues and Practice*,35(2): 6-18.
- [19] Wolf, M. K., Farnsworth, T., & Herman, J. Validity issues in assessing English language learners' language proficiency [J].*Educational Assessment*,13 (2-3): 80-107. doi: 10.1080/10627190802394222.
- [20] Wolf, M. K., Kao, J., Griffin, N., Herman, J. L., Bachman, P. L., Chang, S. M., & Farnsworth, T. Issues in Assessing English Language Learners: English Language Proficiency Measures and Accommodation Uses. Practice Review (Part 2 of 3). CRESST Report 732. National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- [21] Wolf, M. K., Guzman - Orth, D., & Hauck, M. C. Next - Generation Summative English Language Proficiency Assessments for English Learners: Priorities for Policy and Research[J]. ETS Research Report Series,2016(1):1-23.