

## **Ignoring Disadvantaged Students: Caveats of the ELA Regents Exam**

**Jorge Beltran**

*Teachers College, Columbia University*

### **ABSTRACT**

The Comprehensive English Regents Exam, also referred to as the English Language Arts (ELA) Examination, is a widely used measure of student achievement in the K-12 system in New York, yet it is not without criticism. One concern is that the ELA Regents Exam may be contributing to the widening of the gap in graduation rates between English language learners (ELLs) and English-proficient students (NYSED, 2014b). This suggests a threat to the validity of the inferences made based on test scores. When considering the features of test design and the characteristics of the test's diverse target population under an argument-based validity approach (Bachman, 2003, 2005; Kane, 1992, 2002), a mismatch between the intended consequences of test use and the actual effects on the involved stakeholders can be identified. The purpose of this commentary is to analyze possible threats to validity, with issues ranging from domain analysis and construct definition to test misuse from a language assessment perspective and to discuss the impact of the Regents Examination in ELA on newly-arrived ELLs.

### **INTRODUCTION**

Nowadays, standardized tests are one of the most widely used and at the same time strongly criticized measures of student achievement in the K-12 system in the United States. While it is necessary to assess what students are learning to determine the effectiveness of current practices and allocate resources, it is also true that there may be a mismatch between the intended consequences of test use and the actual repercussions on the educational system. This is the case for the Regents Examinations, which have been in use in New York State for more than a century. One of the requirements to graduate from high school is passing the required set of Regents Examinations. However, what was meant to be a way of pushing the system towards better standards has also had a series of negative effects in a number of areas. In particular, the test of Comprehensive English or English Language Arts (ELA) has become an obstacle for many English Language Learners (ELLs), thus creating a gap between graduation rates of ELLs and English-proficient students. These unexpected consequences pose a series of questions regarding the adequacy of the test for this subpopulation as well as a possible threat to the test's validity.

The purpose of this paper is to analyze and discuss the impact of the Regents ELA examination on ELLs as well as consequences of using a test of English as a native language to assess newly-arrived ELLs. First, the historical background of the test will be discussed. Next, the unintended consequences of the test will be considered. Then, a detailed analysis of the ELA

examination will be provided, including possible threats to validity and implications of test misuse for ELLs. Finally, a set of conclusions and recommendations will be provided.

## **The Regents Examinations**

While large-scale testing seems to have become quite prominent in current times, the Regents Examinations have actually been around for a long time. They were first administered to junior high school students as a set of “preliminary” evaluations in November 1865 (NYSED, 2012). In 1878, Regents Examinations for algebra, history, Latin, philosophy, and physical geography began to be used as high school exit exams (NYSED, 2012). These new exams were intended to help certify passing students in an attempt to set instruction standards.

The range of subjects that are tested under the Regents’ test battery has undergone several changes. Subjects come and go depending on modifications in the state core curriculum or governmental mandates. For instance, just in 2011, the remaining tests for foreign languages (French, Spanish, and Italian) were eliminated after a budget cut of \$700,000, relegating the responsibility to account for student proficiency to local academic bodies (Dwyer, 2011).

Nowadays, to graduate from high school, students can either obtain a Regents Diploma or an Advanced Regents Diploma. To obtain a Regents diploma, students entering ninth grade from 2015 onwards have to score a minimum of 65 points in the set of required exams in social studies, math, science, ELA, and one additional subject. To obtain an advanced diploma, students must score 65 or more in three math tests, two science tests, and a foreign language test from the Languages Other Than English (LOTE) tests (NYSED, 2007).

The minimum passing scores have changed slightly throughout the last decade. The tendency has been for cut scores to increase. For example, 55 points was the minimum score to obtain a local diploma in 2005 (NYSED, 2007). The decisions to increase the required test scores, however, have not been accompanied by a solid change in instructional policy that guarantees such growth in student achievement. In fact, it is expected that if the bar is increased for the class of 2022, graduation rates could reach alarming lows (Berris & Murphy, 2014).

In spite of a constant political push for attaining college readiness, one group of students is clearly disadvantaged when being evaluated through this policy: ELLs. Next, the adverse circumstances that ELLs are facing will be discussed, followed by an analysis of the test characteristics that fuel the problems that arise when test-taker traits are not accounted for.

## **The Problem for English Language Learners: Broadening Gaps**

While standardized tests involve many educational stakeholders, there is a particular subpopulation that is directly affected by the lack of thought of the accountability policy: newly-arrived ELLs. According to Menken (2008), in addition to having to get used to a new language and culture, newcomers also have to get used to the testing culture that prevails in U.S. classrooms. Although tests are used as assessment tools worldwide, the extent to which they affect your educational and career opportunities varies greatly from country to country. Therefore, students who are unfamiliar with rigorous learning standards will be at a disadvantage when assessed with a test that targets language skills beyond their language proficiency.

In addition, for a test-taker with limited proficiency of the language that the test is in, any test becomes a language proficiency test (Menken, 2008); therefore posing threats to the validity

of the inferences that can be made by such test scores (Coltrane, 2000). Shohamy (2007) problematized the consequences of assessing language learners through tests in content subjects while ignoring issues of language proficiency. She discussed the impact that this gap in language proficiency may have on performance in subjects like Math. In the ELA Regents, what is being tested is language proficiency within the context of literary analysis, which entails a set of skills and knowledge that is far beyond what many ELLs arriving at or after ninth grade possess.

Given the mismatch of the underlying constructs that a test measures when it is delivered to a population other than the one it was designed for, a series of steps should be taken to account for these issues. Rivera and Stansfield (2001) discussed some types of accommodations that can be provided to ELLs who demonstrate being at a disadvantage for a certain test. These accommodations include additional timing, bilingual dictionaries, and teacher/proctor assistance. However, Rivera and Stansfield also point out possible threats to validity when the implementation of these accommodations is done in such a way that it could also unfairly increase scores of non-ELL students due to assistance beyond linguistic scaffolding.

In the case of the Regents exams, the New York City Department of Education has listed method of presentation, method of response, timing, and setting as possible accommodations for ELLs for up to two years after achieving required proficiency according to the New York State English as a Second Language Achievement Test. School-based teams determine who is eligible to receive special accommodations (NYCDOE, 2014). However, these accommodations may represent a threat to the validity and fairness of the test, given that the administration procedures of the test may vary greatly from one school to another, resulting in construct-irrelevant variance.

As a result of the lack of agreement on the consideration of the particular needs of ELLs, graduation rates for ELLs have reached incredibly low figures since the implementation of the rise in cut scores that establishes 65 as the minimum passing grade. In 2013, only 31.4% of ELLs graduated in comparison to the overall average of 74.9%. Many of these ELLs arrived in the U.S. during high school (NYSED, 2014b). Given that the content-area tests (math, science, and social studies) can be taken in some languages other than English (Spanish, Chinese, Haitian Creole, Korean, or Russian) (NYCDOE, 2015), a big part of this reduced graduation rate could be attributed to the ELA Regents exam, which can only be taken in English.

Considering that in the last ten years there was an increase of 20% in the ELL population in New York State, reflected in an overall 8.9% increase in matriculated ELL students (NYSED, 2014b), the abysmal difference in graduation rates puts into question the adequacy of the graduation requirements policy. In light of these figures, the Board of Regents approved an amendment to section 100.5(7) of the Commissioner's Regulations regarding ELLs. Those ELLs who came to the U.S. in 2010 to begin 9<sup>th</sup> grade or a higher grade will have the option to appeal their score in order to meet graduation requirements when they obtain a score between 55 and 61 and fulfill a set of required stipulations (e.g., 95% rate of attendance and passing all other required Regents exams) (NYSED, 2014b). Although there are many ways in which such measures could be criticized, it represents a first step to establishing a fairer assessment context.

Next, the Comprehensive English Regents exam is analyzed through an argument validity framework in relation to the particular subgroup of ELLs.

## The Validity of the ELA Regents Examination

While the ELA Regents test was not designed as a test of English as a second language, the truth is that it becomes so for about nine percent of test takers who are not native speakers. Therefore, part of the student body will have an inherent disadvantage as they approach the test. In order to understand the issues of test quality, the test will be evaluated in terms of a validity framework.

An argument-based validity framework was developed to be implemented in various assessment contexts. Kane (1992) proposed making interpretive arguments about a chain of inferences needed to establish a test's validity. In the stage of test design, the target domain needs to be considered, particularly regarding the representativeness on the types of performance to be elicited (Kane, Crooks, & Cohen, 1999). The observation inference concerns the adequacy of the scoring in a testing context. The generalization inference refers to the consistency of the scoring and the theoretical invariance of a score, that is, the degree to which a test-taker's score would remain uniform in similar testing conditions (Kane, 1992), which later on was referred to as a *universe score* (Kane et al., 1999). The extrapolation inference addresses the correspondence of a test score with "nontest behavior" (Kane, 1992, p. 529), later on referred to as a *target score* (Kane et al., 1999). Following Mislevy's (2003) use of Toulmin's model for the structure of an inference, Kane (2004) expanded the interpretive argument by considering the decisions based on test scores (therefore implicating test use).

Bachman (2005) elaborated on this and other models (Kane, 1992, 2004; Kane et al., 1999) and proposed a two-part "assessment use argument" (Bachman, 2003; in Bachman, 2005). Bachman's (2005) model consists of a validity argument and a utilization argument. The former connects test takers' performance with interpretations of test scores, while the latter connects these interpretations with test uses and/or decisions. These score-based decisions are validated through evidence supporting the intended warrants and rejecting the possible rebuttals of test use.

Before analyzing each of these steps in validation, let us consider some key elements of the Comprehensive English test. The purpose of the test is to assess student mastery of the core performance indicators from the 2005 English Language Arts Core Curriculum (NYSED, 2014c), Standard 1, which relates to applying language skills for information and understanding, Standard 2, which refers to using language skills for literary response and expression, and Standard 3, which consists of their application to critical analysis and evaluation (NYSED, 2005). While the standards are discussed in terms of all four language skills, the test focuses on listening, reading, and writing. The ELA Regents consists of four sections, a total of 28 items: 25 multiple-choice questions, two short constructed-response questions, and one essay, which are expected to be parallel across forms. The first task is an aural comprehension task, followed by two reading comprehension tasks, and finally the three constructed response items (NYSED 2011).

Looking at what is considered the first step in test validation, domain analysis (Mislevy, Steinberg, & Almond, 2003; in McNamara & Roever, 2006), there are considerable differences between the types of texts selected for the ELA Regents test and those often used in ESL instruction. As determined in the test specifications for the ELA Regents exam, the reading section will include a literary and an informational text, and the integrated skills section (Part 3, Questions 26 and 27) will include two literary elements linked by a theme (NYSED, 2010). Although there is no full description of the types of passages to be included, a review of the ELA tests shows that poems are often used as one of the literary texts for the integrated tasks section.

Poetry is seldom used in ESL examinations given the types of knowledge that it demands from students. Furthermore, this type of text is not directly essential for success in most college or work settings. Most ESL tests focus on informational tests because of the complexity that figurative language represents for students who are yet to learn the literal meanings of the target language. In addition, oftentimes poems tap into what Purpura (2004) calls sociocultural meanings, so test-takers unfamiliar with the cultural context which the literary work depicts will have more difficulty understanding the text.

The observation inference is probably the one that has been studied the most by the department of education. Yearly reports of field tests for new items have been included in technical reports since 2011, displaying mostly psychometrically sound items, and clarifying that only those with reliable properties can be selected to be items in the actual test (NYSED, 2014c). However, these field tests cannot guarantee that the same properties would apply to the actual population of test-takers (e.g., there was not a representative sample to conduct Differential Item Functioning analyses other than on gender). In addition, the scoring of the constructed-response tasks could be problematic. While there are instructions to conduct norming in the test sampler, such as having two raters for the essay question and an available third in case of discrepancies, rater training procedures are designated at the school level (NYSED, 2010). Given the large amount of schools and teachers involved in rating, it is unlikely that this proceeds without inconsistencies, especially given the range of experience and assessment literacy (i.e., knowledge of testing practices) among teachers. While a field test may report great inter-rater agreement among raters, the descriptors of the rubrics for the three constructed-response questions, which are to be used by teachers, seem vague. For instance, as part of the rubric for question 27, in which students have to come up with a paragraph linking two literary texts, the descriptor indicates that for a test-taker to be granted a score of one, he/she either “provides” or “implies an explanation of the literary element or technique” or “has an unclear explanation of the literary element or technique” (NYSED, 2010, p.15) in addition to partial textual evidence, imprecise or inappropriate language, and “errors in convention that may affect meaning” (p.15). In this way, rather than having clear expectations on what a score of 1 may entail, scoring can then become intuitive, given the numerous possibilities of interpretation of the descriptor. For example, an “unclear” explanation could either be approximately correct or completely wrong, but the degree of accuracy of the explanation is not commented upon in the descriptor.

When turning to the generalization inference, which refers to the degree to which a test-taker’s score would remain uniform in similar testing conditions (Kane, 1992), a number of problems arise with a couple tasks. On one hand, there is the first task, which corresponds to a dictation task from which students have to answer 8 multiple choice questions on a passage that the proctor reads out loud twice (NYSED, 2010). A potential problem relates to the consistency in the administration of the test, given that each classroom where the test is being administered will have possible variation attributed to the way in which the proctor reads the passage. In addition, it has been found that the degree of familiarity ELLs have towards a person's speech eases their comprehension. While in many occasions it is their own teachers who proctor the test, in an attempt to prevent bias, there are also situations in which teachers or administrators are asked to proctor classes other than their own (or their own school, which is also a potential disadvantage to the less proficient ELLs).

Although there is a lack of studies on the extrapolation or explanation inferences, a few points should be made. Regarding the explanation inference, it should be noted that while overall the test shows fair to high degrees of internal consistency in its field tests (NYSED, 2014c), there

is no public availability of this information. The broad terms in which the learning standards that the test measures are defined creates an issue of interpretability (Purpura, Brown, & Schoonen, 2015), since it is hard to determine what a student knows through a test score when there is a very loose specification of its scope. Regarding the extrapolation inference, there might be a method effect, considering that there is no justification as to why some test-takers may pass the Regents in ELA Common Core test but not the traditional ELA exam (Burris & Murphy, 2014).

When considering the utilization inference, while it can be argued with the warrant of utility (Bachman, 2015) that a low ELA score for ELLs represents their poor skills on literary analysis in English, this raises an issue of justice (McNamara & Ryan, 2014), since this expectation is unrealistic for many newly-arrived students. The gap in graduation rates is a clear example of how this mandate affects ELLs with unintended consequences (see the type two rebuttal in Bachman's 2005 model). Moreover, these high-stakes decisions have resulted in an unfortunate washback effect; so that many schools (and teachers) seem to be focusing on attaining higher scores on the Regents (also considering that their scores are part of teacher evaluations). All this has an effect on the quality of instruction, since learning objectives are narrowed to "passing a test," often without consideration of the several other skills that are needed for success in higher education which are not embodied in the Regents tests.

Although what has been discussed presents a difficult panorama of issues for ELLs in New York and questions the Regents test itself, actions are being taken to overcome issues of validity and fairness.

## **Modifications Underway**

As of 2014, the new Regents in ELA Common Core Test is being implemented across schools in New York State as the Comprehensive English test is being phased out (NYSED, 2010). The new Regents in ELA Common Core Test assesses the reading and writing learning standards of the New York State P-12 Common Core Learning Standards for English Language Arts & Literacy, which are to be implemented in curricula and assessments based on the assumption that "a successful integration of the standards will provide students with the fluency, comprehension, analytic, and communication skills necessary to be on track for college and career readiness" (NYSED, 2014a, p. 2). The test comprises three reading and two writing tasks, including writing an argumentative essay instead of a completely artificial task. In spite of this step forward, there have also been criticisms. According to Burris and Murphy (2014), this new test is actually "too difficult," with texts almost three times longer than the traditional test (a 4,000 word difference), and increased difficulty in vocabulary, theme, and syntax. Overall, however, Burris and Murphy (2014) argue that the writing section has been weighed so heavily that a student with a regular performance in writing would pass even with a very low score in reading. There needs to be a careful analysis on how the reading and writing sections should be weighed in relation to the learning standards that are being assessed. In this way, while attempts to address issues of validity and fairness are being undertaken by the current ELA Common Core Test, it is possible that they are moving in the wrong direction, so the validity of this test and alignment with curricular learning standards needs to be carefully examined.

## CONCLUSION

In this paper, the validity and impact of test use of the Comprehensive English Regents Exam has been analyzed, and an argument has been made for the inadequate and unfair practice of conditioning graduation for ELLs when such a test represents a challenge for this particular subpopulation of students. In view of this, a few things could be done to reduce gaps in graduation rates, for example, establishing an alternative graduation path for ELLs, whether this entails designing an alternate test form, standardizing ELL accommodations, or setting an empirically based set of cut scores. An alternate test form for ELLs would be viable as long as considerations of the differences between taking an ELA test as a native English speaker and as an ELL are taken. Standardized accommodations would probably be easier to implement, but the type of accommodation that works best in this context should be determined empirically. Finally, cut scores for ELLs could be set based on a study of the degree to which ELLs can master the learning standards that an ELA test targets within the scope of what is possible within the course of study in high school.

The rationale for adopting one of these measures should be grounded on the development of feasible learning objectives for ELLs. Requesting a score in a test that is mediated by language proficiency does not seem to be a fair measure for exit requirements, particularly when language learning is different from other content areas, and time is necessary for acquisition. This appears especially needed when considering the difference in constructs when the English is a test-takers second language.

Making changes in policy and/or curricula in consideration of the unique needs of certain student populations could prevent tests like the Regents ELA exam from functioning as gatekeepers, and instead act as gate-openers (Bachman and Purpura, 2008). After all, the Regents exams are meant to be powerful and objective tools to provide an opportunity to compare outcome achievement and quality of instruction across the state, not to keep students from ever graduating high school.

## REFERENCES

- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2(1), 1–34.
- Bachman, L. F., & Purpura, J. E. (2008). Language assessments: Gate-keepers or gate-openers? In B. Spolsky & F. M. Hult (Eds.), *The Handbook of Educational Linguistics* (pp. 456-468). Oxford, UK: Blackwell Publishing.
- Burris, C. & Murphy, J. (2014). A disturbing look at the Common Core tests in New York. *The Washington Post*. Retrieved on December 10, 2015 from: <https://www.washingtonpost.com/news/answer-sheet/wp/2014/06/27/a-disturbing-look-at-common-core-tests-in-new-york/>
- Dwyer, J. (2011, May 31). A step back for learning languages. *The New York Times*, pp. A20. Retrieved from [http://www.nytimes.com/2011/06/01/nyregion/a-step-back-for-learning-languages.html?\\_r=0](http://www.nytimes.com/2011/06/01/nyregion/a-step-back-for-learning-languages.html?_r=0)
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527-535.

- Kane, M., Crooks, T., & Cohen, A. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice*, 18(2), 5–17.
- McNamara, T., & Roever, C. (2006). Validity and the social dimension of language testing. *Language Learning*, 56 (Supplement 2), 9-42.
- McNamara, T. & Ryan, K. (2011). Fairness vs justice in language testing: The place of English literacy in the Australian Citizenship Test. *Language Assessment Quarterly*, 8(2), 161–78.
- Menken, A. (2008). *Language learners left behind: Standardized testing as language policy*. Tonawanda, NY: Multilingual Matters.
- Mislevy, R. J. (2003). Argument substance and argument structure in educational assessment. *Law, Probability and Risk*, 2(4), 237–258.
- NYCDOE (2014) *Testing accommodations-guidance for students and families*. Retrieved on December 13, 2015 from: [http://schools.nyc.gov/NR/ronlyres/D0479BAC-2FA2-43C2-BB64-073FCEE7A99C/0/TestingAccommodations\\_Families\\_FINAL.pdf](http://schools.nyc.gov/NR/ronlyres/D0479BAC-2FA2-43C2-BB64-073FCEE7A99C/0/TestingAccommodations_Families_FINAL.pdf)
- NYCDOE (2015). *Yearly testing*. Retrieved on December 12, 2015, from: <http://schools.nyc.gov/Accountability/resources/testing/default.htm#CT%20ELLS>
- NYSED (2005). *English language arts core curriculum (Prekindergarten-Grade 12)*. Retrieved on May 1, 2016 from: <http://www.p12.nysed.gov/ciai/ela/elacore.pdf>
- NYSED (2007). *Summary of NYSED regulation 100.5 and Chancellor's regulation A.501*. Retrieved on December 10, 2015 from: <http://schools.nyc.gov/NR/ronlyres/215FF06B-DCA3-442B-89DF-18E674DC867E/0/Acpolicygened.pdf>
- NYSED (2010). *Comprehensive English sampler, Spring 2010*. Retrieved on December 11, 2015 from: <http://www.nysedregents.org/comprehensiveenglish/english-sampler-2010.pdf>
- NYSED (2012). *History of Regents Examinations*. Retrieved on December 10<sup>th</sup>, 2015 from: <http://www.p12.nysed.gov/assessment/hsgen/archive/rehistory.htm>
- NYSED (2014a). *Educator guide to the 2014 Regents Examination in ELA (Common Core)*. Retrieved on May 1, 2016 from: [https://www.engageny.org/file/17336/download/english\\_language\\_arts\\_common\\_core\\_test\\_guide.pdf](https://www.engageny.org/file/17336/download/english_language_arts_common_core_test_guide.pdf)
- NYSED (2014b). *Graduation requirements*, October 2014. Retrieved from: <http://www.p12.nysed.gov/biling/docs/RegentsItemGraduationRequirementOctober2014.pdf>
- NYSED (2014c). *NYS Regents Examination in Comprehensive English: 2014 field test analysis, equating procedure, and scaling of operational test forms technical report*. Retrieved on March 31, 2016 from: <http://www.p12.nysed.gov/assessment/reports/2014/compeng-tr14.pdf>
- Purpura, J. E. (2004) *Assessing grammar*. New York, NY: Cambridge University Press.
- Purpura, J. E., Brown, J. D., Schoonen, R. (2015). Improving the validity of quantitative measures in applied linguistics research. *Language Learning*, 65(Supplement 1), 36-73.
- Rivera, C. & Stratisfield, C. (2001). *Test accommodations for LEP students*. Retrieved on December 10, 2015 from: <http://www.ericdigests.org/2002-2/lep.htm>
- Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing*, 18(4), 373-391.

Jorge Beltran is a student of the Ed. D. in Applied Linguistics at Teachers College Columbia University in the language assessment track. His current research interests include learning-oriented assessment, assessment of speaking, construct validation, and performance-based assessment. He can be reached at: [jl2262@tc.columbia.edu](mailto:jl2262@tc.columbia.edu)