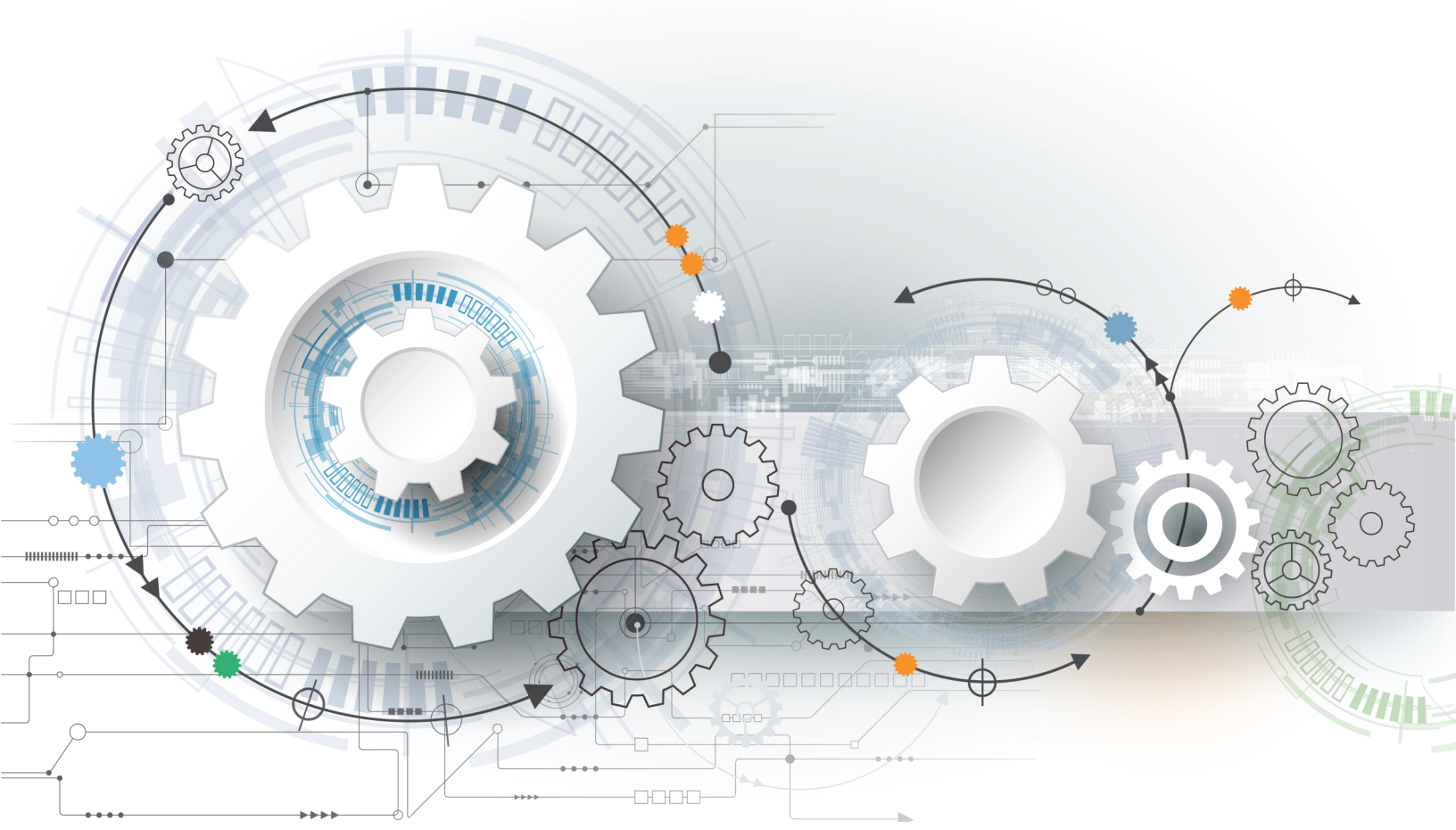
Working Paper
2018-1

Testing Supports for English Learners:

A Literature Review and Preliminary ACT Research Findings

JOANN MOORE, PHD, CHI-YU HUANG, PHD, NOOREE HUH, PHD,
TIANLI LI, PHD, AND WAYNE CAMARA, PHD



ACT.org/research

ABOUT THE AUTHORS

Joann Moore is a research scientist in Validity and Efficacy Research specializing in prediction of secondary and postsecondary outcomes from academic and non-cognitive factors.

Chi-Yu Huang is a principal psychometrician in Psychometrics Research specializing in research and anaslysis in educational psychometrics, data forensics, and accommodations.

Noo-Ree Huh is a senior psychometrician in Psychometrics Research specializing in research and analysis in educational psychometrics, data forensics, and accommodations.

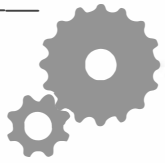
Tianli Li is a senior psychometrician in Assessment Transformation specializing in psychometric research and supports for large scale testing programs.

Wayne Camara is ACT's Horace Mann Chair. His research background spans the ACT continuum of solutions. Before he came to ACT, he held positions at the American Psychological Association and College Board, where he served as vice president of research and development.

ACKNOWLEDGEMENTS

The authors would like to thank Kurt Burkum, Krista Mattern, and Tracy Wilkinson for their comments on an earlier draft of this report.

ACT WORKING PAPER SERIES



ACT working papers document preliminary research. The papers are intended to promote discussion and feedback before formal publication. The research does not necessarily reflect the views of ACT.

Abstract

In the fall of 2017, ACT began providing a limited number of supports (also known as accommodations) to English Learner (EL) students in the US taking the ACT® test.[1] The goal of the supports is to remove construct-irrelevant variance in students' scores related to limited English proficiency and allow students to more accurately demonstrate their true proficiency on the construct being measured. This paper presents a summary of previous research on the impact of providing supports to EL students and a summary of a preliminary analysis conducted by ACT based on data from states who administered the ACT statewide, captured students' EL status, and provided accommodations to EL students. ACT's future research agenda for providing validity evidence for this initiative is also discussed.

[1]ACT (2016, November 14). *ACT to provide supports for English learners on the ACT test* [press release]. Retrieved from http://www.act.org/content/act/en/newsroom/act-to-provide-supports-for-english-learners-on-the-act-test-.html

Background

Testing supports or accommodations are modifications to a test or test administration conditions meant to reduce construct-irrelevant variance and allow for more meaningful measurement of a given construct so that a student's scores more accurately reflect their knowledge or ability on the measured construct. For example, a student with a visual impairment may be provided with a screen reader or a braille version of a test so that they can access the test content. For students for whom English is not their first or best language, assessment of subjects other than the English language may result in scores that are not reflective of the student's mastery of that content, particularly if the assessment requires a lot of reading (Abedi, Leon, & Mirocha, 2003; Noble, Rosebery, Suarez, Warren, & O'Connor, 2014).

For many years, ACT has had policies in place allowing students with documented disabilities to take the ACT using relevant ACT-approved accommodations. However, prior to the fall of 2017, ACT did not allow accommodations based on limited language proficiency alone. In recent years, the ACT has been increasingly adopted for use in state accountability programs. For this use, ACT allows states to administer the test using accommodations and supports beyond those that are approved by ACT, including supports based on limited language proficiency. However, the scores resulting from these special testing conditions are not college reportable; only scores resulting from standard administration conditions or ACT-allowed accommodations are college reportable.

The rules for the Every Student Succeeds Act (ESSA) has stipulated that all groups should receive the same benefits from assessments when a district opts to use a nationally recognized high school assessment in lieu of the state assessment for accountability. This has been widely interpreted to recommend that ACT or SAT scores resulting from state/district

testing and used for federal accountability should allow use for the same purposes as scores from other groups (e.g., admissions, scholarships, and placement). In other words, all students testing in this context should receive college-reportable scores.

Beginning in the fall of 2017, ACT began providing college-reportable scores for US EL students who are approved to take the ACT using one or more of four relevant supports:

- Additional time on the test (time-and-a-half)
- Use of an approved word-to-word bilingual glossary (containing no word definitions)
- Test instructions provided in the student's native language (including Spanish and a limited number of other languages initially)
- Testing in a non-distracting environment (i.e., in a separate room)

These supports were reviewed and approved as appropriate by content experts at ACT. The accommodations were also reviewed by a panel of external experts in research, measurement, higher education, English learners, state and federal policy, and civil rights. Research on the impact of extended time for ELs has shown mixed results, but research on the other three supports has generally indicated that they have a low threat of impacting construct validity (Abedi, Courtney & Leon, 2003; Abedi & Ewers, 2013; Abedi, Hofstetter, Baker, & Lord, 2001; Acosta, Rivera, Willner, 2008; Cawthon, Ho, Patel, Potvin, Trundt, 2009; Francis, Rivera, Lesaux, Kieffer, & Rivera, 2006; Kieffer, Rivera & Francis, 2012; Lovett, 2011; Pennock-Roman & Rivera, 2011).

Other next-generation assessments being used for accountability have also adopted policies for the use of supports by EL students. The Partnership for Assessment of Readiness for College and Careers (PARCC) allows EL students with proper documentation to use select

supports on their assessments, including extended time and bilingual/dual language word-to-word dictionaries on all assessments, as well as the use of speech-to-text and human scribes and readers during the mathematics assessment only (Partnership for Assessment of Readiness for College and Careers, 2017). The Smarter Balanced Assessment Consortium (SBAC) allows EL students to use universal tools and designated supports, including pop-up English glossaries and translated test directions and glossaries, but states that "no accommodations are available for ELLs on the Smarter Balanced assessments; accommodations are only available to students with disabilities and ELLs with disabilities" (Smarter Balanced Assessment Consortium, 2017, p. 33). The National Assessment of Educational Progress (NAEP) allows EL students to use a bilingual dictionary without definitions (National Center for Education Statistics, 2015). In January 2017, the College Board began allowing testing accommodations for EL students taking the SAT as part of state testing contracts, which included translated test instructions in a limited number of languages and an approved word-to-word bilingual glossary. They are currently conducting research on extra time and indicated they would include this accommodation if score validity is not impacted.

Challenges in Studying English Learners

There are a large number of challenges in evaluating the impact of accommodations on performance of English learners. First, unlike a permanent disability, EL status should be temporary, assuming that the student is receiving adequate support and instruction. Thus, EL students should receive accommodations based on their *current* English proficiency status, which may not be the same accommodations they received in previous years.

Another major challenge in studying the performance of EL students is that they are a vastly diverse, heterogeneous group of students, differing with respect to many characteristics, such as:

- Native language
- Level of proficiency in their native language (and may have different levels of proficiency in reading, writing, speaking, or listening)
- Level of proficiency in English (and may have different levels of proficiency in reading, writing, speaking, or listening)
- Level of academic proficiency in non-English subject areas
- Number of years of English instruction they have had (in the US or elsewhere)
- Number of years of instruction in their native language
- Interruptions in instruction
- Number of years the student has been in the US
- Number of years the student has been in an EL program
- Criteria used to determine English proficiency level
- Criteria for exiting EL services
- English-language proficiency of their parents or guardians
- Amount and type of English language support they receive at home
- Amount and type of English language support they receive at school
- If the student also has a disability
- Socio-economic status (SES)

Because of the diversity of the EL population, some accommodations may work better for some students than others, and some may not be appropriate for some students. The specific

accommodations provided to an individual student should take into account their specific needs. Historically, accommodations provided to EL students may have been simply adopted from accommodations created for students with disabilities and may not be appropriate for EL students (Abedi, 2009; Acosta, Rivera, & Willner, 2008) or may be ineffective (Abedi, Hofstetter, & Lord, 2004; Solano-Flores, Wang, Kachchaf, Soltero-Gonzalez, & Nguyen-Le, 2014). For example, a Spanish translation would not be appropriate for examinees who do not understand Spanish, and even within the Spanish language, there are dialects that vary depending on the students' native country (Sireci & Faulkner-Bond, 2015; Solano-Flores & Li, 2009).

EL students face additional challenges beyond limited English proficiency. For instance, EL students are more likely to come from families with lower income and lower levels of parent education (Abedi, 2002; Abedi, Leon, & Mirocha, 2003) and may have less opportunity to learn tested material than non-EL students (Abedi, 2003; Herman & Abedi, 2004). Parent education levels and socio-economic status are highly related to academic performance (Abedi, Leon, & Mirocha, 2003), meaning that EL students from low income families may be doubly disadvantaged as compared to their peers. Similarly, EL students are more likely to be taught by teachers who are less experienced or lack appropriate teaching credentials (Abedi & Gandara, 2006; National Research Council, 2000), and teachers may not be able to identify appropriate accommodations for their students (Fuchs, Fuchs, Eaton, Hamlett, Binkley, & Crouch, 2000; Fuchs, Fuchs, Eaton, Hamlett, & Karns, 2001; Helwig & Tindal, 2003; Kopriva, Emick, Hipolito-Delgado, & Cameron, 2007; Koran & Kopriva, 2006; Koran & Kopriva, 2008; Weston, 2003). Different accommodations may also have different impacts on EL student performance depending on the student's background (Abedi, 2003; Abedi, Hofstetter, & Lord, 2004).

An additional challenge in studying EL students is that while ESSA provides a federal definition of an English learner (US Department of Education, 2016), the criteria used to identify EL students, determine who is eligible for accommodations, and determine which accommodations to provide are established locally and often differ by school, district, or state (Abedi, 2001; Abedi, 2008a; Abedi, 2008b; Abedi, Hofstetter, & Lord, 2004; Linquanti, Cook, Bailey, & MacDonald, 2016). The most frequent methods used for identifying EL students include home language surveys, parental reporting, teacher observations, student records, teacher interviews, and referrals, all of which may also vary across states (Bailey & Carroll, 2015; Linquanti, Cook, Bailey, & MacDonald, 2016; Young & King, 2008). Different districts may have different academic standards for English language proficiency and may use different English proficiency assessments; therefore, students with the same level of English proficiency may differ in their English proficiency status depending on where they live (Abedi, Hofstetter, & Lord, 2004; Young & King, 2008). For these reasons, consistent classification of EL students continues to be a major limitation of research findings (Abedi, 2004; Camara, 2009; Garcia, 1991; Hafner, 2001; Reschly, 1996). Additionally, the types of services and the duration of services EL students receive (e.g., bilingual instruction, pull-out services, language supports) may differ across districts for the same student (Kindler, 2002) and may depend on state policy or resources (Abedi, 2003).

ACT determines EL support eligibility based on documentation provided by a school test coordinator or test accommodations coordinator. Documentation may include an EL Plan, Individualized Education Program (IEP), an official support or accommodations plan, EL proficiency assessment results, and/or confirmation of eligibility or participation in an EL program (ACT, 2017b). However, if the criteria for identifying and classifying EL students are

inconsistent across schools, then it is difficult to ensure that ACT is offering supports consistently to the populations who need them, and it is also difficult to study the impact of the supports on specific segments of the EL population (for example, which accommodations are appropriate for students at various levels of English proficiency) if they are not consistently identified or classified.

In contrast, the process used to identify students with disabilities is based on documentation of a specific diagnosis by a qualified physician or other qualified professional whose credentials are appropriate to the disability (ACT, 2017a). Documentation may include an IEP, Section 504 plan, or official accommodations plan, and detailed documentation is required, including a statement of the diagnosed impairment and descriptions of the presenting problems, the student's developmental history, limitations resulting from the impairment, rationale for the recommended accommodations and how the accommodations alleviate the impact of the impairment, professional credentials of the person providing the evaluation, and the assessments used to inform the diagnosis. Ideally, the identification of both English learners and students with disabilities should be standardized and consistently applied. To the extent that the process for ELs is more variable, it makes it more difficult to ensure that students are appropriately identified, given the services that they need, and provided appropriate assessment supports or accommodations.

Previous Research

A review of research that has been conducted on the impact of providing accommodations to EL students illustrates the challenges with studying such a diverse population; namely, that it is difficult to find consistent effects when comparing studies across accommodation types, subject areas, grade levels, and different populations of students. Several

meta-analyses have found mixed results of the effectiveness of various accommodations, with some studies showing a positive effect, and others showing little or no effect, or even a negative effect, depending on the students' grade level, native language, level of English proficiency and native language proficiency, or other factors (Chiu & Pearson, 1999; Francis et al., 2006; Kieffer, Rivera, & Francis, 2012; Li & Suen, 2012; Pennock-Roman & Rivera, 2011; Sireci, Li, & Scarpati, 2003).

Ideally, an accommodation should benefit the students who need it but should not benefit students who do not need it, a concept known as the interaction hypothesis or differential boost hypothesis (Lane & Levanthal, 2015; Sireci, Scarpati, & Li, 2005). While some accommodations are not likely to benefit non-EL students (for example, a word-to-word Spanish translation glossary), some accommodations such as extra time may benefit all students. Sireci, Li, & Scarpati (2003) evaluated the interaction hypothesis in a review of 40 studies pertaining to the effects of accommodations on students with disabilities and English learners. They concluded that in many studies, extra time improved performance of all students but tended to benefit students with disabilities and English learners to a greater extent. They proposed a revision to the interaction hypothesis such that if an accommodation benefits students with disabilities and EL students significantly more than it benefits fluent English speakers without disabilities, then the accommodation may indeed reduce construct-irrelevant variance in these populations. They also noted the inconsistency of findings across studies due to different accommodation types, different research designs (e.g., experimental, quasi-experimental, ex post facto), and heterogeneity of student groups, and found that accommodations for EL students tended to be more beneficial when extra time was allowed.

Chiu and Pearson's (1999) meta-analysis included 30 studies of accommodations for special education and EL students across K-12 and postsecondary populations. They found an overall weighted mean effect size of 0.16 standard deviations when comparing accommodation and no accommodation conditions of special education and EL students. General education students also benefitted from accommodations, but to a smaller degree (0.06 standard deviations). The authors also noted a wide amount of heterogeneity across studies, such that in some cases students benefitted from accommodation, sometimes there was no effect, and sometimes there was a negative effect.

Francis et al. (2006) conducted a meta-analysis including 11 studies that employed random assignment of EL and non-EL students to accommodated and non-accommodated conditions. Of seven accommodation types (simplified English, English dictionary or glossary, bilingual dictionary or glossary, extra time, Spanish language test, dual language items, and dual language booklet), only one was found to have a significant positive effect on EL performance: English language dictionaries and glossaries, with a mean effect size of 0.146, corresponding to a 10% to 25% reduction in the performance gap between EL and non-EL students (Kieffer, Lesaux, Rivera, & Francis, 2009).

Pennock-Roman and Rivera (2011) focused on 14 studies that employed either random assignment or counterbalanced repeated measures to investigate the impact of accommodations on EL and non-EL students, taking into account English proficiency level and restricted time versus extra time. Eleven types of accommodations were considered (plain/simplified English, bilingual glossary, Spanish version, extra time, English dictionary/glossary, pop-up English glossary, dual language, picture dictionary, pop-up bilingual glossary, English read aloud, and small group administration). The authors found that Spanish versions of tests were the most

effective accommodation for Spanish-speaking EL students with low levels of English proficiency, with effect sizes of 0.95 and 1.45 in two separate studies, and plain/simplified English was the most effective accommodation for EL students with high intermediate levels of English proficiency, with an effect size of 0.57. They also found that in general, effect sizes were larger for studies with extra time or untimed conditions as compared to studies with restricted time limits and concluded that extra time is often necessary to allow students to make use of the accommodations.

Kieffer, Rivera, and Francis (2012) extended the Francis et al. (2006) meta-analysis, including 20 studies employing experimental, quasi-experimental, or counterbalanced, repeated measures designs and nine accommodation types (simplified English, English dictionary or glossary, bilingual dictionary or glossary, extra time, native language test, dual language booklet, reading the test aloud, and dual language booklets read aloud). They found significant positive effects for three of the accommodations: simplified English, English dictionaries or glossaries, and extra time. Simplified English was associated with a 9% to 19% reduction in the achievement gap between EL and non-EL students, English glossaries or dictionaries was found to have an 11% to 21% reduction in the achievement gap, and extra time provided a 15% to 31% reduction in the achievement gap.

Li and Suen (2012) conducted a meta-analysis of 19 experimental and quasi-experimental studies using a hierarchical linear modeling (HLM) approach to take into account ethnicity, grade level, subject area, English proficiency, and accommodation type. The authors found an overall effect size of 0.157 standard deviations. Of the predictor variables included in the HLM model, only English proficiency was significant, suggesting that the accommodations provided a larger benefit to students with lower English proficiency.

Some of the mixed results found in the literature may be due to interactions between the type of accommodation provided and student characteristics; for example, EL students with little formal instruction or limited proficiency who cannot read in their native language would not be likely to benefit from written translations or a bilingual glossary or dictionary (Abedi, Hofstetter, & Lord, 2004; Francis et al., 2006). How EL students are grouped for analysis may also have an impact on whether a benefit is found. Ercikan, Roth, Simon, Sandilands, and Lyons-Thomas (2014) found that different items were flagged for differential item functioning (DIF) depending on whether students were considered a linguistic minority or linguistic majority in their school. Oliveri, Ercikan, and Zumbo (2014) used simulated data to examine the effects of linguistic heterogeneity within EL groups and found decreased detection of DIF and increased false negatives when heterogeneous groups of students were included as the comparison group.

Limited previous research evaluating the internal structure of tests for EL students tested with accommodations also shows mixed results. Young, Cho, Ling, Cline, Steinberg, and Stone (2008) found that the unidimensionality of a science assessment for 8th-grade students was supported for EL students who were provided access to bilingual glossaries; results were mixed for other grade levels and other assessments. Using an experimental design, Abedi, Lord, Hofstetter, and Baker (2000) found high internal consistency for EL and non-EL students taking math and reading assessments, with slightly lower coefficients for EL students. Steinberg, Cline, Ling, Cook, and Tognatta (2009) conducted DIF and factor analyses for 4th- and 8th-grade students taking an English Language Arts (ELA) assessment. Four groups were compared: EL students without disabilities, non-EL students without disabilities, non-EL students who were deaf or hard of hearing, and EL students who were deaf or hard of hearing. They found few items

with moderate to large DIF favoring the non-disabled, non-EL group and found that a single factor solution adequately fit the data for all four groups in both grades.

Students' English proficiency level may confound the extent to which they benefit from accommodations. Li and Suen's (2012) meta-analysis found that students with lower levels of English proficiency benefited more across a variety of accommodations than students with higher levels of English proficiency. Alternatively, Albus, Bielinski, Thurlow, and Liu (2001) found that EL students with higher levels of English reading proficiency using a monolingual English dictionary showed significant score gains, whereas students with lower English proficiency and non-EL students did not benefit from the use of a dictionary.

Linguistic simplification may benefit some EL students (Abedi, Courtney, Mirocha, Leon, & Goldberg, 2005), but it may not be effective if it doesn't take into account language heterogeneity (Solano-Flores, 2006). Linguistic simplification may also benefit non-EL students. Abedi and Lord (2001) found that reduced linguistic complexity benefited both EL and non-EL students, but performance gains were higher for EL, low ability, and low income students.

Students classified as EL tend to score lower, on average, than students who are English proficient (Abedi, 2002; Abedi, 2009; Abedi, 2014; Abedi, Lord, Boscardin, & Miyoshi, 2001), and in observational data, EL students who receive accommodations tend to score lower than EL students who do not receive accommodations (Young et al., 2008), suggesting that EL students who receive accommodations may have lower levels of English proficiency than those who do not receive accommodations. The performance gap tends to be larger in English Language Arts than mathematics (Abedi, 2002), and the gap increases as the linguistic complexity of the assessment increases (Abedi, Leon, & Mirocha, 2003). Providing accommodations to EL

students may increase student performance but does not completely eliminate the performance gap (Anderson, Liu, Swierzbin, Thurlow, and Bielinski, 2000).

Abedi (2002; 2009) proposed a Disparity Index (DI) to quantify the differential boost or impact among two groups. The DI is calculated by subtracting the mean of the reference group from the mean of the focal group and dividing by the mean of the focal group. This value is then multiplied by 100 to convert it to a percentage that distinguishes the performance disparity among various groups. A negative value results when performance by the focal group is lower than that of the referent group. Abedi (2009) reported that EL students in 3rd-grade reading underperformed non-EL students by 53%, whereas EL students with disabilities underperformed non-EL students by over 200%.

As when testing students with disabilities, combinations of accommodations are also typically used for EL students; for example, many accommodations are accompanied by extra time, making it more difficult to disentangle the effects of a given accommodation or to obtain adequate sample sizes to investigate the effects of some accommodations or combinations of accommodations (Francis, et al, 2006; Rivera & Collum, 2004). In fact, small samples are often an issue in evaluating performance of special tested populations and the effects of accommodations for these students. In many environments, only a small proportion of examinees are English learners (Pennock-Roman & Rivera, 2011). This issue is exacerbated when a state or school district may have EL students from more than 100 different linguistic backgrounds (O'Conner, Abedi, & Tung, 2012a; O'Conner, Abedi, & Tung, 2012b; Sireci & Faulkner-Bond, 2015). This makes it nearly impossible to obtain adequate samples to evaluate the impacts of specific accommodations for a specific language, a specific level of English proficiency, or for a specific level of proficiency in reading, writing, speaking, or listening in English. Much of the

research is based on heterogeneous groups of EL students, grouping together EL students with different native languages, with different levels of English proficiency, and across different types of accommodations.

Kopriva, Emick, Hipolito-Delgado, and Cameron (2007) noted that most research of the impact of accommodations on EL performance involves providing the same accommodations or sets of accommodations to groups of EL students, regardless of the students' specific needs, and that teachers tend to provide the same set of accommodations to all of their students despite being able to recognize that the students have diverse needs. They found that EL students given appropriate accommodations benefitted more than EL students given inappropriate or incomplete accommodations, and inappropriate or incomplete accommodations were no more beneficial than no accommodations at all.

Accommodations Research Relevant to EL Supports Offered on the ACT

Additional time.

ACT allows extra time up to time-and-a-half for EL students approved for this support. Additional time is the most common accommodation provided to English learners and students with disabilities (Abedi, Courtney, & Leon, 2003; Abedi, Hofstetter, Baker, & Lord, 2001; Acosta, Rivera, & Willner, 2008; Cawthon, Ho, Patel, Potvin, & Trundt, 2009; Kieffer, Rivera & Francis, 2012; Lovett, 2011; Pennock-Roman & Rivera, 2011). English learners require more time to comprehend the language of the assessment than fluent English speakers (Francis et al., 2006). They also require additional time to make use of accommodations, for example, to look up words in a glossary (Pennock-Roman & Rivera, 2011).

Abedi, Lord, and Plummer (1997) found higher omit rates on a math test for students who indicated that they spoke a language other than English at home. Schnipke and Pashley

(1997) found that students whose primary language was not English responded more slowly than English speakers during the first half of a reasoning test and exhibited more rapid guessing behaviors during the second half of the assessment, suggesting that the test was more speeded for non-English speakers. Zenisky and Baldwin (2006) found that EL students spent significantly more time completing math and reading tests than non-EL students. Crotts (2013) found that EL students had significantly longer response times than non-EL students; this was true both with and without accommodations. Roohr and Sireci (2017) found that EL students had significantly longer response times in completing test items than did non-EL students, and accommodated items took significantly longer to complete than non-accommodated items.

Several studies have found that providing extra time provided more of a benefit in performance for EL students than for non-EL students, partially supporting the interaction hypothesis (Abedi, 2009; Abedi, Courtney, & Leon, 2003). Other studies have found that both EL and non-EL students benefited from extra time (Abedi, Lord, Hofstetter, & Baker, 2000; Hafner, 2001; Lovett, 2011). Such findings are a central issue with differential boost; determination of whether additional time benefits EL students more than non-EL students and whether the resulting scores are more reflective of EL students' actual ability are key issues in determining the validity of this accommodation. Additionally, other accommodations may not be effective if students do not have sufficient time to make use of the accommodation (Pennock-Roman & Rivera, 2011). Extra time itself is not going to help EL students whose English proficiency it too low for them to be able to understand the test content (Rivera & Stansfield, 1998).

While there have been studies on the impact of providing extra time to EL students, there is not an empirical basis to provide guidelines for the appropriate amount of extra time that

should be provided for specific students. For example, the ideal amount of extra time to be provided may be different for students at different levels of English proficiency, for different combinations of accommodations (e.g., a bilingual glossary or dictionary), or for EL students who also have a disability.

Much of the research on accommodations has either been conducted on tests which are untimed or generously timed, whereas admissions tests are much more aggressively timed. The speededness of a test could moderate the impact accommodations have on EL students' performance and the validity of scores, as well as score comparability, and need to be examined in context. A few studies have examined speededness and extra time on college admissions tests, mostly in the context of students with disabilities.

Ziomek and Andrews (1998) found that extra time on the ACT improved the performance of students with disabilities who first tested under standard time and retested with extra time by an average of 3.2 points. Students with disabilities who tested twice with extra time saw score gains of 0.9 points, which is similar to that of students without disabilities who tested twice under standard time (0.7). Camara, Copeland, and Rothschild (1998) found that extra time on the SAT improved the performance of students with disabilities; students with disabilities who first tested under standard time and retested with extra time saw average score gains that were three times higher than the gains of students without disabilities who tested twice with standard time. A limitation of these two studies is that they used operational test data and were unable to examine the impact of extra time for non-disabled students.

Hill (1984) found that learning disabled students took significantly longer to complete an untimed ACT test (243.5 minutes) than non-learning disabled students (184.7 minutes; standard administration was 160 minutes). Learning disabled students saw significant score gains; total

raw score averages were 98 under standard time and 139 when untimed. Non-learning disabled students saw much smaller gains; total raw score averages were 141 under standard time and 149 when untimed. Ragosta and Wendler (1992) found that 1.5 to 2 times the testing time was required by students with disabilities to ensure that approximately equal percentages of disabled and non-disabled students completed each section of the SAT. Mandinach, Bridgeman, Cahalan-Laitusis, and Trapani (2005) found that time-and-a-half on the SAT benefitted students both with and without disabilities; gains were higher for students with disabilities but were nonsignificant due to small sample sizes. Benefits of extended time were greater on the math section than on the verbal section, and low-ability students did not benefit from extra time.

Bridgeman, Trapani, and Curley (2004) used shortened SAT test forms and found that more time per question resulted in minimal score gains on the verbal test (no more than 10 points on a 200-800 scale) and somewhat larger gains on the math test for students scoring in the 510-700 range (30 points on a 200-800 scale). Extra time did not benefit low-ability students in math. They did not find any differential benefits of extra time for students whose best language was not English. Bridgeman and Cline (2007), in a SAT field trial, found that extra time provided no benefits on a reading test, gains of 0.08-0.09 standard deviations on a math test, and gains of 0.16 standard deviations on a multiple choice writing test.

Bilingual glossary.

Bilingual glossaries or dictionaries are another common accommodation provided to EL students (Abedi, 2014; Thurlow, Larson, Lazarus, Shyyan, & Christensen, 2017; Wolf, Kao, Griffin, Herman, Bachman, Chang, & Farnsworth, 2008). ACT allows the use of an approved word-to-word bilingual glossary containing no word definitions for EL students approved for this support. Prior research conducted on the impact of providing English learners with a bilingual

glossary or dictionary suggests that they do benefit EL students (Francis et al., 2006; Kieffer, Lesaux, Rivera, & Francis, 2009; Kieffer, Rivera & Francis, 2012). A meta-analysis by Francis et al. (2006) found that the only accommodation with an overall positive effect was the use of English language dictionaries or glossaries. However, much of the literature has focused on English or bilingual glossaries or dictionaries that include word definitions; few studies have investigated the effectiveness of word-to-word translations containing limited or no definitions (Abedi, Courtney, & Leon, 2003; Abedi, Courtney, Mirocha, Leon, & Goldberg, 2005; Kopriva, Emick, Hipolito-Delgado, & Cameron, 2007; Rivera & Collum, 2004).

Abedi, Courtney, and Leon (2003) found that a pop-up glossary containing brief definitions or synonyms of unfamiliar, non-content-related terms had a positive impact on math performance of EL students without impacting performance of non-EL students. Abedi, Courtney, Mirocha, Leon, and Goldberg (2005) also found that bilingual glossaries containing translations but no definitions resulted in improved performance relative to 4th- and 8th-grade EL students tested without accommodations on a science test. Kopriva, Emick, Hipolito-Delgado, and Cameron (2007) included a bilingual glossary (a pop-up containing a Spanish translation) as one of several accommodations in a study comparing the effects of appropriate, incomplete or not recommended, and no accommodations provided to 3rd- and 4th-grade EL students on a math assessment. While they did not specifically test for statistical significance of the bilingual glossary, EL students who received a bilingual glossary had higher scores than EL students who did not receive any accommodations, averaging approximately one point higher on a 30 item test.

Pennock-Roman and Rivera's (2011) meta-analysis found that bilingual dictionaries were more likely to benefit EL students with low English proficiency, whereas English dictionaries

were more useful for EL students with intermediate levels of English proficiency. Abedi (2009) found that a pop-up glossary was more effective than a customized printed dictionary, likely due to unfamiliarity with dictionary use; only 9% of 4th-grade students and 14% of 8th-grade students in the study reported that they had used a dictionary in the past. Abedi, Courtney, and Leon (2003) also found that a pop-up glossary was more effective than a customized dictionary, and students reported greater use of the glossary than the dictionary.

Rivera and Collum (2004) noted that one of the difficulties in comparing studies of the use of glossaries or dictionaries are inconsistencies in how the terms are defined and applied across studies. They define a dictionary as a compilation of words that are defined generally, whereas a glossary provides definitions that are "customized for a particular context and audience" (Rivera & Collum, 2004, p. 8). Additionally, dictionaries or glossaries might be customized to simplify language specifically for EL students, they may or may not provide examples of word usage or synonyms, they may only include words that are relevant to the assessment and have been reviewed to ensure that they do not contain content that would alter the construct of interest, or they may contain translations. Research reports varied in the level of detail about whether glossaries or dictionaries contained word-to-word translation, synonyms, simple definitions, or full definitions. Rivera and Stansfield (1998) noted that bilingual glossaries that do not provide definitions and customized glossaries that only include relevant content are expected to be the most useful without endangering the construct of interest.

Abedi, Courtney, and Leon (2003) noted that dictionaries containing definitions could provide an unfair advantage to examinees if they provide clues or answers to test items. They found that providing customized glossaries containing only terms (with definitions) that are not part of the tested content resulted in mixed results, benefitting some students but not others, and

that there was a considerable amount of effort required in producing the customized dictionaries, making this a difficult accommodation to provide at scale. Acosta, Rivera, and Willner (2008) also recommended against providing dictionaries that contain definitions because they could provide an unfair advantage. Instead, they recommended creating lists of commercial English or word-to-word dual language dictionaries approved for use with specific assessments.

Abedi, Courtney, Mirocha, Leon, and Goldberg (2005) suggest that students who use a bilingual glossary on a regular basis in the classroom may benefit more from using them on an assessment than students who are not familiar with their use. They found that providing an English or bilingual glossary to 4th-grade EL students resulted in higher scores on a science test than EL students who tested under standard conditions, while non-EL students' performance was not impacted. However, translations may not be helpful if instruction has only been in English (Abedi, Lord, Boscardin, & Myoshi, 2001; Abedi, Lord, & Hofstetter, 1998).

Francis et al.'s (2006) meta-analysis did not find a positive impact of bilingual glossaries or dictionaries on EL student performance. They suggest that the effects of bilingual glossaries may depend on the populations being tested or other contextual factors. Kieffer, Rivera, and Francis (2012) reached a similar conclusion, noting the heterogeneity of effect sizes of bilingual glossaries across studies. Koran and Kopriva (2008) also suggest that bilingual glossaries will only be effective for students with at least some literacy in their home language.

Test instructions in native language.

ACT provides written test instructions in the student's native language (12 languages,[2] including Spanish, at this time) for EL students approved for this support. Providing instructions in students' native language is a commonly used accommodation (Lazarin, 2006).

Abedi and Ewers (2013) concluded that reading aloud test directions in students' native language does not pose much risk of altering the construct of interest. Hafner (2001) investigated the impact of providing "extended oral presentation," including simplifying test directions, re-reading directions, and reading directions in students' native language. Sample sizes were small and an overall positive effect of accommodation was found, but the effects of extra time and extended oral presentation were not examined separately. To our knowledge, there has not been research conducted investigating the effect of providing written translated test instructions to EL students, but it is reasonable to assume that this accommodation would only benefit students who are proficient in reading in their native language.

Test in separate room.

Testing in a separate room is a common accommodation, often used in conjunction with other accommodations to provide a non-distracting environment. Because this accommodation is often confounded with other accommodations, it is unclear whether it has an impact on the performance of EL students on its own (Abedi, Courtney, & Leon, 2003; Abedi, Lord, Boscardin, & Miyoshi, 2001; Acosta, Rivera, & Willner, 2008).

[2] The 12 languages for which written test instructions were available in fall 2017 were Arabic, Chinese (Simplified), Chinese (Traditional), French, German, Haitian Creole, Korean, Russian, Somali, Spanish, Tagalog, and Vietnamese.

Best Practices

Much has been written in recent years regarding best practices and recommendations for providing appropriate testing accommodations for EL students (Abedi & Ewers, 2013; Acosta, Rivera, & Willner, 2008; Francis et al., 2006; Kieffer, Rivera, & Francis, 2012; Rivera, Acosta, Rivera, & Willner, 2008; Wolf, Herman, & Dietel, 2010). In general, accommodations should be tailored to the specific needs of the student (Francis et al., 2006; Kopriva, Emick, Hipolito-Delgado, & Cameron, 2007), and an accommodation should benefit students who need the accommodation, but not benefit students who do not need the accommodation (Abedi, Hofstetter, & Lord, 2004; Shepard, Taylor, & Betebenner, 1998). Accommodations for EL students should take into account students' English proficiency level as well as their level of proficiency in their native language (Acosta, Rivera, & Willner, 2008). Testing accommodations should also be familiar to the students receiving them (Abedi, Courtney, Mirocha, Leon, & Goldberg, 2005; Acosta, Rivera, & Willner, 2008). Research supports the use of simplified English for assessments where English comprehension is not the construct of interest, and English dictionaries or glossaries have shown a positive impact (Kieffer, Rivera, & Francis, 2012). Accommodations may also be subject-area specific (Francis et al., 2006). For example, reading the items aloud may be appropriate for a test measuring computation skills but would be inappropriate for a test measuring reading comprehension.

Summary of Prior Research

There is no single study or methodological design that is likely to sufficiently address the two long-standing questions which remain when introducing accommodations or test modifications to a particular group: (1) whether accommodations sufficiently minimize the impact of test-taker attributes irrelevant to the construct; and (2) the extent to which

accommodations introduce construct irrelevant factors which distort score validity (Camara, 2009). A comprehensive review of such questions includes the type of correlational studies examining the differential boost hypotheses as described earlier in this paper, factor analysis studies, research on differential item functioning, and research on differential prediction which contrasts groups (e.g., EL) in various accommodations conditions (e.g., extended time, no extended time, multiple accommodations). In general, the findings across correlational studies of the impact of providing accommodations to EL students have been mixed. In large part, this is to be expected because of the diversity of EL students, the different ways in which a given accommodation may be provided to EL students, the extent to which the student actually uses the accommodation, and different impacts of a given accommodation on student performance given each student's specific needs. Linguistic modification, dictionaries or glossaries, and extra time have shown some effectiveness (Kieffer et al., 2009; Sireci et al., 2003, Pennock-Roman & Rivera, 2011; Steinberg, Cline, Ling, Cook, and Tognatta, 2009), and there is some evidence in favor of the use of the four supports that will be provided to EL students on the ACT, suggesting that they may have a positive benefit to EL students while not violating the constructs measured by the assessment.

Predicting College Performance of English Learners

Because the primary purpose of the ACT is its use as a college admissions assessment, it is vital that ACT scores are accurate predictors of student performance in college. While additional research is needed to examine the relationship between ACT scores and college performance for EL students, several studies of other college admissions assessments have found that for students whose best language was not English, testing without accommodations resulted in underprediction of college performance.

A 1994 study by the College Board compared first year college performance of students whose self-reported best language was English or another language and found that the first year grade point average (FYGPA) of students whose best language was a language other than English tended to be underpredicted by the SAT-Verbal but not the SAT-Mathematical (Ramist, Lewis, & McCamley-Jenkins, 1994). They also found underprediction of course grades in quantitative courses and slight overprediction in English courses. A 2008 study found similar results; FYGPA of students whose best language was a language other than English tended to be underpredicted by the SAT Critical Reading and Writing assessments, but not by the SAT Mathematics assessment (Mattern, Patterson, Shaw, Kobrin, & Barbuti, 2008).

A 2012 study found that the SAT Critical Reading, Mathematics, and Writing tests underpredicted first year college mathematics course grades, and SAT Critical Reading and Writing tests underpredicted first year college English course grades for students who indicated that their best language was not English (Mattern, Patterson, & Kobrin, 2012). Conversely, Zwick and Schlemer (2004) found evidence that the SAT overpredicted first year college GPA for Latino and Asian students whose first language was not English. Another study (Zwick & Sklar, 2005) found little evidence that SAT scores underpredicted FYGPA for Hispanic students reporting Spanish as their first language. Note that this study defined the linguistic minority group as students for whom Spanish was their *first* language rather than students for whom a language other than English was their *best* language. This study also was based on small samples; there were 267 students in the regression analyses whose first language was Spanish. One limitation of these studies was that the samples of students for whom English was not their first or best language were small, ranging from 100 in the 2012 study to 1,718 for the 2008 study.

Preliminary ACT Research Findings

While further research is needed to evaluate the impact of providing ACT-approved supports to EL students on the ACT, preliminary analyses have been conducted using data from ACT statewide testing, including EL students who tested with state-approved accommodations where the resulting scores were not college-reportable. The following analyses are based on data from two state testing programs using the ACT for 11th grade accountability reporting across five years (2010-2014), with permission from the states to use student data. Table 1 contains the counts and percentages of EL and non-EL students in the two data samples, including EL students taking the ACT without accommodations and EL students taking the ACT with state-allowed accommodations in the two states. The data sample from State 1 includes 539,083 students, and the data sample from State 2 includes 645,410 students. Two percent of students in each state were English learners. Less than one percent of EL students in each state took the ACT with state-allowed accommodations.

Table 1. Non-EL and EL Students in Two States

	State 1		State 2		Total	
Testing Group	Frequency	Percent	Frequency	Percent	Frequency	Percent
Non-EL, No Accommodations	525,894	98%	632,189	98%	1,158,083	98%
EL, No Accommodations	12,867	2%	10,385	2%	23,252	2%
EL, State-Allowed Accommodations	322	0.1%	2,836	0.4%	3,158	0.3%
EL Total	13,189	2%	13,221	2%	26,410	2%
Total	539,083		645,410		1,184,493	

Because the focus of this research is on EL students, students who took the ACT with accommodations due to having a disability were excluded from these analysis. However, it is possible that some of the EL students who tested with state-allowed accommodations might also have had one or more diagnosed disabilities, and it is possible that the accommodation was

provided due to limited English proficiency, the disability, or to a combination of limited English proficiency and disability. There was not sufficient information in the data samples to definitively determine which EL students also had a disability, which is a limitation of this study. Another limitation of this study is that we could not reliably discern which accommodations the EL students received; it is unknown whether the accommodations provided were appropriate to students' specific needs or the extent to which they are comparable to the supports allowed by ACT. Additionally, due to small sample sizes and incomplete information on the type of accommodation(s) provided, the results could not be disaggregated by accommodation type. Due to small sample sizes, the two states' data were combined for the remainder of the analyses presented in this brief.

Examinee ACT Score Performance

The ACT contains four sections: English, mathematics, reading, and science (with an optional writing section, excluded from these analyses) and is reported on a scale from 1-36 (with the exception of the writing section, which is reported on a scale of 2-12). Table 2 contains the average ACT scores and Disparity Index of non-EL students who took the ACT without accommodations, EL students who took the ACT without accommodations, and EL students who took the ACT with state-allowed accommodations. On average, EL students had substantially lower ACT scores than non-EL students (mean ACT Composite = 20.4), and among EL students, those who tested with state-allowed accommodations had lower scores (mean ACT Composite = 13.5) than those who tested without accommodations (mean ACT Composite = 14.8). EL students in both testing groups tended to have lower scores in English and reading relative to their performance in math and science.

The Disparity Index is a measure of the percent of difference in performance of a focal group (in this case, the two EL student testing groups) relative to a reference group (non-EL students). A negative value indicates that the focal group performed worse relative to the reference group. As can be seen in Table 2, both groups of EL students performed worse, on average, than non-EL students across all subject areas, and the difference was more pronounced in English and reading. The disparity in performance was higher for EL students testing with accommodations than EL students testing without accommodations; ACT Composite scores of EL students testing with accommodations were 51.1% lower than the ACT Composite scores of non-EL students, and ACT Composite scores of EL students who tested without accommodations were 37.9% lower than average ACT Composite scores of non-EL students.

Table 2. Average ACT Scores and Disparity Index (DI) of EL and Non-EL Students from Two States

Testing Group	Subject	Mean	Standard Deviation	Disparity Index
Non-EL, No Accommodations (N= 1,158,083)	English	19.8	6.4	
	Math	20.5	5.2	
	Reading	20.2	6.0	
	Science	20.5	5.2	
	Composite	20.4	5.2	
EL, No Accommodations (N= 23,252)	English	12.5	4.2	-58.4
	Math	16.5	3.5	-23.9
	Reading	13.9	3.6	-45.5
	Science	15.6	3.9	-31.0
	Composite	14.8	3.1	-37.9
EL, State-Allowed Accommodations (N= 3,158)	English	10.5	3.2	-88.2
	Math	15.9	3.2	-28.5
	Reading	12.7	2.8	-59.5
	Science	14.6	3.5	-39.6
	Composite	13.5	2.4	-51.1

To examine whether state-allowed accommodations impacted performance, score changes of students who took the ACT more than once were evaluated. Students who took the ACT as part of a national test administration within one year after the ACT state test were

included in the analysis. EL students did not receive accommodations as part of the national test administrations, allowing us to examine differences in scores for students who tested with accommodations on one occasion and without accommodations on another. State test records were matched to national test records using first name, last name, date of birth, and state of residence.

Because the ACT, when used for accountability, is typically offered during the spring of students' 11th-grade year and most students first take the ACT between the spring of their 11th-grade year and the fall of their 12th-grade year, most of the students in the analysis sample took the state test first. This is the case for both EL and non-EL students. Due to small sample sizes, students who took the national test first are omitted, and only students who took the state test first are included in subsequent analyses to ensure that order effects are not impacting the results. In general, students' scores tend to increase slightly when they take the ACT a second time, likely due to both familiarity with the test and learning that occurs between the two test events (Andrews & Ziomek, 1998; Camara & Allen, 2017; Harmston & Crouse, 2016).

Twenty percent of the EL students were found to have taken the ACT a second time within a year after taking the state test compared to 30% of non-EL students. EL students who tested without accommodations were more likely to test a second time within a year (20%) compared to EL students who tested with accommodations (18%).

Because the sample size of non-EL students was much larger than that of EL students in the analysis, and because non-EL students had higher performance on the ACT than EL students, a second group of non-EL students was created to be more comparable in sample size and student ability to the EL students. The matched group included 10,000 randomly selected non-EL students whose ACT Composite scores on the state test were matched to the score

distributions of the EL students on the state test. Five random samples were created, and the average difference scores between the state and national tests from these five samplings were used. Because students with lower ACT scores are more likely to increase their scores upon retest than students with higher ACT scores (Andrews & Ziomek, 1998; Camara & Allen, 2017), the matched group provides a more comparable baseline to the full non-EL group when evaluating score gains of EL students.

Table 3 contains average ACT scores, Disparity Indices (DI), score gains, and effect sizes of score gains for students who took the ACT on a national test administration within one year after taking the ACT during statewide testing. Compared to the full sample of state-tested students included in Table 2, students who retested during a national administration tended to be higher performing. An examination of average scores and DI show that on both the state and national tests, EL students tended to score substantially lower than non-EL students, and EL students who tested with accommodations had lower scores and higher DI than EL students who tested without accommodations. The disparities in performance were highest in English and reading for both groups of EL students, particularly the English performance of EL students who tested with accommodations, who performed 80% lower than non-EL students on the state test and 78% lower on the national test.

Also of note is that for EL students who tested without accommodations, the disparities in performance were lower on the national test than on the state test (suggesting a possible motivation effect on the state test), but for EL students who tested with accommodations on the state test, the DI was lower on the state test for all but the English test, suggesting that the accommodations did indeed reduce the disparity in performance for other subject areas besides English.

Average score gains of the full non-EL group ranged from 0.26 in math to 0.83 in reading, which were comparable to score gains of the EL students who tested without accommodations, which ranged from 0.27 in math to 0.82 in reading. Score gains of EL students who tested with state-allowed accommodations were much smaller, ranging from -0.15 in math to 0.46 in English. These smaller gains for EL students who tested with accommodations on the state test and without accommodations on the national test is what we would expect if the accommodations had a positive impact on their state test performance. This suggests that the accommodations did benefit the students. Score gains of the non-EL sample that was matched to be more similar in performance to EL students were larger than the gains of the other testing groups, ranging from 0.29 in math to 1.27 in reading. Because students with lower ACT scores are more likely to increase their scores upon retest than students with higher ACT scores (Andrews & Ziomek, 1998; Camara & Allen, 2017), the lower score gains of the EL students testing without accommodations relative to the matched group could indicate that they are less able to fully access the test content. It is also possible that the larger gains of the matched group can be explained in part by regression to the mean, a phenomenon in which samples selected on the basis of extreme scores tend to show less extreme scores upon retest due to the influence of measurement error in the first set of scores.

Table 3. ACT Performance, Disparity (DI), and Score Gains for Students Taking State Test First and Retesting on a National Test Administration

		State Test		National Test		Difference (National – State)	Effect Size (Cohen's *d*)
Testing Group	Subject	Mean (SD)	DI	Mean (SD)	DI	Mean (SD)	
	English	22.0 (5.9)		22.6 (5.9)		0.62 (2.75)	0.11
	Math	22.4 (5.2)		22.7 (5.3)		0.26 (2.33)	0.05
Non-EL, No	Reading	21.7 (5.7)		22.5 (5.8)		0.83 (3.58)	0.15
Accommodations	Science	22.0 (4.7)		22.5 (4.8)		0.46 (3.23)	0.10
(N = 353,115)	Composite	22.2 (4.8)		22.7 (4.9)		0.55 (1.76)	0.11
	English	14.9 (4.8)	-48.0	15.7 (4.7)	-44.6	0.79 (2.83)	0.17
	Math	18.6 (4.5)	-20.6	18.8 (4.8)	-20.3	0.27 (2.17)	0.06
EL, No	Reading	15.3 (4.0)	-42.1	16.1 (4.2)	-40.1	0.82 (3.47)	0.20
Accommodations	Science	17.5 (4.0)	-25.7	18.2 (3.9)	-23.6	0.68 (3.54)	0.17
(N = 4,720)	Composite	16.7 (3.6)	-32.8	17.3 (3.7)	-31.0	0.65 (1.86)	0.18
	English	12.2 (3.6)	-80.2	12.7 (3.5)	-78.4	0.46 (3.10)	0.13
	Math	18.0 (4.2)	-24.5	17.9 (4.0)	-27.0	-0.15 (2.36)	-0.04
EL, State-Allowed	Reading	13.5 (3.4)	-60.0	13.9 (3.4)	-62.3	0.32 (3.96)	0.09
Accommodations	Science	16.5 (3.9)	-33.7	16.7 (3.6)	-35.0	0.19 (3.98)	0.05
(N = 560)	Composite	15.1 (2.9)	-47.0	15.3 (2.9)	-48.1	0.26 (2.20)	0.09
	English	15.5 (4.7)	-42.0	16.6 (4.7)	-36.6	1.06 (2.83)	0.23
Non-EL, No	Math	17.3 (3.4)	-29.2	17.6 (3.7)	-28.6	0.29 (2.06)	0.08
Accommodations,	Reading	16.0 (4.1)	-35.3	17.3 (4.4)	-30.1	1.27 (3.42)	0.30
Matched Group	Science	17.0 (4.1)	-29.4	18.1 (4.0)	-24.4	1.06 (3.56)	0.26
(N = 7,877)	Composite	16.6 (3.5)	-33.7	17.5 (3.6)	-29.6	0.95 (1.81)	0.27

Overall, it appears that the state-allowed accommodations may have helped mitigate the language barrier and allowed EL students to increase their performance on the ACT. However, because the sample size of the group of EL students with state-allowed accommodations was small, and many factors (such as students' test-taking motivation, time elapsed between state and national tests, learning interventions, and accommodation types) were not considered in the study, the results from this study were insufficient to provide definitive evidence of the impact of state-allowed accommodations on the scores of EL students.

High School Grades

Because the ACT is primarily a measure of college readiness, a key piece of validity evidence is to demonstrate that providing accommodations to EL students results in accurate prediction of college performance. If EL student performance on the ACT is negatively impacted by limited English proficiency, and if accommodations remove the construct-irrelevant variance and allow students to more accurately demonstrate their true achievement level, then we would anticipate that providing the supports would result in more accurate prediction of college performance than EL students testing without supports. At this time, there are no relevant studies available using admissions test scores to predict college grades for the three groups of interest in this study (Non-EL, No Accommodations; EL, No Accommodations; and EL, State-Allowed Accommodations). As noted earlier, until fall of 2017 accommodations have only been available for EL students testing as part of a state administration of the ACT, and those accommodations have been provided to only a small number of students resulting in non-college reportable scores. However, even as such accommodations are being introduced in fall 2017 for EL students nationally, obtaining sufficient data from EL students with and without accommodations will be a challenge. Validity studies require outcome data from colleges and universities, and the

number of EL students in each condition enrolling in even large universities will make it difficult to obtain sufficiently large and representative samples required to examine differential prediction and bias.

High school GPA can be used as a surrogate for college GPA and provide some insight into the issues of differential prediction and bias. Students provide their high school coursework taken and grades as part of the ACT registration process. Previous ACT research has found that self-reported high school grades are highly accurate (Sanchez & Buddin, 2015). Other research has found moderate median correlations amongst high school grade point average (HSGPA) derived from self-reported grades, ACT Composite scores, and first year college grade point average across 192 institutions (Sawyer, 2010). In Sawyer's (2010) study, the correlation between ACT Composite and HSGPA ($r = 0.44$) was slightly higher but comparable to the correlation between ACT Composite and first year college GPA ($r = 0.41$), and the correlation between high school GPA and college GPA was slightly higher ($r = 0.48$), with high school GPA explaining about one-quarter of the variance in college GPA ($R^2 = 0.23$). High school and college GPA are both influenced by factors such as student achievement, conscientiousness, and other student-level factors which are likely to remain stable as students transition from high school to college, as well as factors that may differ between high school and college, such as courses taken, grading standards, and other student factors related to the high school to college transition. While there are limitations of using high school grades as a surrogate for college grades, doing so can help us understand whether ACT scores of EL students are negatively impacted by limited English proficiency.

The data sample considered in the analysis of relationships between ACT scores and student grades is a subset of the sample presented in Table 3, consisting of all students in the two

states who administered the ACT as part of statewide testing, took the ACT again under standard administration conditions during a national test date within a year after taking the state test, and reported their high school grades when they registered to take the national test. This allows us to compare relationships between ACT scores and student grades for EL students who tested with accommodations on one occasion and tested without accommodations on another occasion. Because some students failed to provide their high school grades, sample sizes are smaller, and the samples may not be representative of the overall testing groups.

It should also be noted that the correlations between ACT Composite and HSGPA ($r = 0.58$ for the non-EL students) are much higher than that found in Sawyer (2010; $r = 0.44$); this is because the correlations in the Sawyer (2010) paper are based on students who enrolled in college, whereas the analyses presented here were not limited to students who enrolled in college. The group in the Sawyer (2010) paper also had higher mean scores (ACT Composite = 22.6) than the groups presented here.

Table 4 contains ACT score performance for the students in two states who took the ACT as part of statewide testing, retested on a national test date within a year after taking the state test, and reported their high school grades. Table 5 contains their average HSGPA, both overall and by subject area. Overall, 92% of students provided high school grades. Non-EL students were most likely to provide high school grades (92%), compared to EL students who tested without accommodations (91%) and EL students who tested with state-allowed accommodations (81%). A comparison of Table 3 and Table 4 shows that students who reported their high school grades tended to score slightly higher on the ACT (between 0.1 and 0.4 score points higher) than students who did not report their grades.

Table 4. ACT Performance, DI, and Score Gains of Students with Self-Reported HSGPA on National Test

		State Test		National Test		Difference (National – State)	Effect Size (Cohen's *d*)
Testing Group	Subject	Mean (SD)	DI	Mean (SD)	DI	Mean (SD)	
Non-EL, No Accommodations (N = 322,136)	English	22.2 (5.9)		22.8 (5.8)		0.62 (2.73)	0.11
	Math	22.5 (5.2)		22.8 (5.3)		0.27 (2.32)	0.05
	Reading	21.8 (5.7)		22.7 (5.8)		0.84 (3.57)	0.15
	Science	22.1 (4.6)		22.6 (4.8)		0.47 (3.21)	0.10
	Composite	22.3 (4.8)		22.8 (4.9)		0.56 (1.74)	0.12
EL, No Accommodations (N = 4,349)	English	15.0 (4.7)	-48.1	15.8 (4.7)	-44.7	0.78 (2.81)	0.17
	Math	18.7 (4.5)	-20.6	19.0 (4.9)	-20.3	0.27 (2.17)	0.06
	Reading	15.3 (3.9)	-42.8	16.1 (4.2)	-41.1	0.78 (3.42)	0.19
	Science	17.6 (4.0)	-25.8	18.3 (3.9)	-23.7	0.67 (3.50)	0.17
	Composite	16.8 (3.6)	-33.0	17.4 (3.7)	-31.4	0.64 (1.82)	0.18
EL, State-Allowed Accommodations (N = 424)	English	12.4 (3.8)	-78.3	12.9 (3.6)	-77.1	0.37 (3.19)	0.10
	Math	18.4 (4.4)	-22.5	18.1 (4.0)	-26.2	-0.34 (2.48)	-0.08
	Reading	13.7 (3.6)	-58.8	13.9 (3.5)	-62.5	0.18 (4.11)	0.05
	Science	16.9 (3.9)	-31.3	16.9 (3.6)	-33.9	0.04 (3.92)	0.01
	Composite	15.3 (3.0)	-45.4	15.5 (2.9)	-46.9	0.19 (2.32)	0.06

Table 5 shows that the average overall HSGPA of EL students who tested without accommodations (mean = 3.26) was slightly higher than the HSGPA of EL students who tested with accommodations (mean = 3.24); both groups had lower HSGPAs than non-EL students (mean = 3.38). These patterns held across subject area grades as well.

Table 5. Average HSGPA of Students with Self-Reported HSGPA on National Test

Testing Group	HSGPA Subject	Mean	Standard Deviation
Non-EL, No Accommodations (N = 322,136)	English	3.44	0.59
	Math	3.28	0.68
	Social Studies	3.49	0.58
	Science	3.32	0.64
	Overall	3.38	0.53
EL, No Accommodations (N = 4,349)	English	3.30	0.64
	Math	3.25	0.73
	Social Studies	3.32	0.67
	Science	3.18	0.71
	Overall	3.26	0.57
EL, State-Allowed Accommodations (N = 424)	English	3.29	0.69
	Math	3.26	0.71
	Social Studies	3.25	0.71
	Science	3.18	0.70
	Overall	3.24	0.59

Table 6 contains intercorrelations amongst students' ACT scores on the state test and HSGPA for each of the three testing groups described in Table 4. Correlations between ACT Composite scores and HSGPA were highest for non-EL students (r = 0.58), lower for EL students who tested without accommodations (r = 0.44), and lowest for EL students who tested with accommodations (r = 0.36). This trend held across subject areas but was most pronounced for English and reading. The correlations between HSGPA and ACT reading scores were markedly higher for non-EL students (r = 0.48) than EL students testing without accommodations (r = 0.28) and EL students testing with state-allowed accommodations (r =

0.15). Intercorrelations between subject area scores on the ACT were also higher for non-EL students than for EL students, particularly the correlations between ACT reading and ACT math scores.

Table 6. Intercorrelations amongst ACT State Test Scores and HSGPA for Retested Students

Testing Group	ACT Subject	HSGPA	ACT English	ACT Math	ACT Reading	ACT Science
	English	0.54				
	Math	0.55	0.75			
Non-EL, No	Reading	0.48	0.79	0.66		
Accommodations	Science	0.52	0.74	0.78	0.71	
(N = 322,136)	Composite	0.58	0.92	0.88	0.89	0.89
	English	0.35				
	Math	0.44	0.56			
EL, No	Reading	0.28	0.70	0.39		
Accommodations	Science	0.38	0.62	0.64	0.53	
(N = 4,349)	Composite	0.44	0.88	0.79	0.78	0.84
	English	0.29				
	Math	0.34	0.49			
EL, State-Allowed	Reading	0.15*	0.50	0.24		
Accommodations	Science	0.27	0.48	0.60	0.40	
(N = 424)	Composite	0.36	0.80	0.79	0.65	0.80

Note: All correlations significant at $p < 0.0001$ unless asterisked.
* Significant at $p < 0.01$.

Table 7 contains intercorrelations amongst students' ACT scores on the national test administration and HSGPA for the same students whose statewide test performance is presented in Table 6. Similar to the state test results, correlations between ACT Composite scores and HSGPA were highest for non-EL students' ($r = 0.58$), lower for EL students who tested without accommodations ($r = 0.42$), and lowest for EL students who tested with accommodations ($r = 0.31$). Comparing the correlations of Table 6 and Table 7, differences in correlations were very small for the two groups who tested without accommodations. However, when comparing the correlations for EL students who took the state test with accommodations and the national test without accommodations, we can see that the relationship between test scores and high school

grades was stronger when the students tested with accommodations. The correlation between ACT Composite scores and HSGPA was higher for EL students who took the ACT with accommodations during the state test (r = 0.36) than the correlation for those same students when they took the ACT on a national test date under standard administration conditions (r = 0.31). In particular, EL students' English and reading scores from the state test with accommodations were more predictive of HSGPA than their English and reading scores from the national administration. These results provide support that providing accommodations to EL students may result in ACT scores that more accurately reflect student performance.

Table 7. Intercorrelations amongst ACT National Test Scores and HSGPA for Retested Students

Testing Group	ACT Subject	HSGPA	ACT English	ACT Math	ACT Reading	ACT Science
	English	0.54				
	Math	0.56	0.76			
Non-EL, No	Reading	0.47	0.79	0.67		
Accommodations	Science	0.51	0.74	0.79	0.72	
(N = 322,136)	Composite	0.58	0.92	0.89	0.89	0.89
	English	0.36				
	Math	0.43	0.59			
EL, No	Reading	0.27	0.70	0.43		
Accommodations	Science	0.35	0.63	0.67	0.57	
(N = 4,349)	Composite	0.42	0.87	0.81	0.80	0.84
	English	0.25				
EL, State-	Math	0.35	0.49			
Allowed	Reading	0.12*	0.49	0.35		
Accommodations	Science	0.27	0.50	0.59	0.49	
(N = 424)	Composite	0.31	0.78	0.79	0.73	0.82

Note: All correlations significant at $p < 0.0001$ unless asterisked.
* Significant at $p < 0.05$.

Table 8 contains correlations between ACT scores on the state administration and on the national administration for the same groups of students presented in Table 6 and Table 7. This provides a measure of test-retest reliability for students who tested with and without accommodations. Within subject area correlations (in bold) were highest for non-EL students

(ranging from 0.77 for science to 0.90 for math), lower for EL students who tested without accommodations on the state test (ranging from 0.61 in science to 0.90 in math), and lowest for EL students who tested with accommodations on the state test and without accommodations on the national test (ranging from 0.32 in reading to 0.83 in math). Of note are the much lower correlations between ACT reading scores and other subject area tests for EL students who tested with supports on the state test and without supports on the national test. We would expect lower test-retest correlations for EL students who received accommodations if the accommodation had a positive impact on performance.

Table 8. Intercorrelations between State and National ACT Test Administrations

	National Test	State Test				
Testing Group	ACT Subject	English	Math	Reading	Science	Composite
Non-EL, No Accommodations (N= 386,612)	English	**0.89**	0.74	0.78	0.73	0.88
	Math	0.74	**0.90**	0.66	0.77	0.85
	Reading	0.77	0.65	**0.81**	0.70	0.82
	Science	0.72	0.77	0.69	**0.77**	0.82
	Composite	0.87	0.85	0.82	0.82	**0.94**
EL, No Accommodations (N= 4,979)	English	**0.82**	0.57	0.67	0.61	0.81
	Math	0.54	**0.90**	0.38	0.62	0.74
	Reading	0.65	0.40	**0.65**	0.50	0.66
	Science	0.58	0.65	0.48	**0.61**	0.70
	Composite	0.77	0.76	0.65	0.70	**0.87**
EL, State-Allowed Accommodations (N= 452)	English	**0.63**	0.45	0.41	0.45	0.65
	Math	0.43	**0.83**	0.17*	0.54	0.67
	Reading	0.30	0.22	**0.32**	0.29	0.36
	Science	0.43	0.50	0.28	**0.46**	0.55
	Composite	0.56	0.65	0.36	0.56	**0.71**

Note: All correlations significant at $p < 0.0001$ unless asterisked.
* $p < 0.001$.

Predicting High School Grades from ACT Scores

To further investigate the relationship between HSGPA and ACT scores, a regression analysis was conducted predicting HSGPA from ACT Composite scores on this same subset of students (as described in Table 4). While this is technically not a prediction model because high

school grades temporally precede ACT scores, we are using HSGPA as a proxy for first year college GPA, which is our actual outcome of interest. Analysis of the residuals (actual GPA – predicted GPA) can be used to assess whether ACT scores are underpredicting the performance of EL students, which we would expect to see if lack of English proficiency is indeed impacting performance on the ACT.

Table 9 contains mean actual HSGPA, predicted HSGPA, and residual scores (actual HSGPA – predicted HSGPA, in bold) for the three testing groups' performance on the state and national tests. Positive residuals indicate underprediction, and negative residuals indicate overprediction. Results were similar for the state and national tests. EL students' high school grades were underpredicted compared to non-EL students by about one-quarter of a grade point for EL students who tested without accommodations and by about one-third of a grade point for EL students who tested with accommodations. It may seem counterintuitive that the accommodated group was underpredicted more than the non-accommodated group. However, recall that there are several limitations of the data sample. We don't know what accommodation(s) students in the accommodated group received, whether they were appropriate, or whether the accommodation were provided due to disability. Additionally, we do not know the level of English proficiency of the students in either testing group, and we also know that states differ with respect to how EL students are identified and what services they receive in school. The accommodated group also had the lowest average ACT scores and GPAs of the three groups; students receiving state-allowed accommodations may have a lower level of English proficiency such that even with accommodations they may be unable to demonstrate their true ability on the ACT.

Table 9. Mean Residuals for Actual – Predicted HSGPA for the State and National Tests

		State Test		National Test	
Testing Group	HSGPA	Mean	Standard Deviation	Mean	Standard Deviation
Non-EL, No Accommodations (N= 386,612)	Actual	3.38	0.53	3.38	0.53
	Predicted	3.38	0.31	3.38	0.30
	Residual	**0.00**	0.43	**0.00**	0.43
EL, No Accommodations (N= 4,979)	Actual	3.26	0.57	3.26	0.57
	Predicted	3.03	0.23	3.04	0.23
	Residual	**0.23**	0.52	**0.22**	0.52
EL, State-Allowed Accommodations (N= 452)	Actual	3.24	0.59	3.24	0.59
	Predicted	2.94	0.19	2.93	0.18
	Residual	**0.30**	0.56	**0.31**	0.56

State model intercept $\beta_0 = 1.97$, coefficient for ACT Composite $\beta_1 = 0.06$, model $R^2 = 0.33$.
National model intercept $\beta_0 = 1.97$, coefficient for ACT Composite $\beta_1 = 0.06$, model $R^2 = 0.33$.

Previous studies of non-EL students have found subject-specific underprediction (Ramist, Lewis, & McCamley-Jenkins, 1994; Mattern, Patterson, Shaw, Kobrin, & Barbuti, 2008; Mattern, Patterson, & Kobrin, 2012). Therefore, mean residuals for actual – predicted HSGPA were also calculated for each ACT subject test and each subject-specific HSGPA. Results are presented in Table 10 for the state test and Table 11 for the national test. Mean residuals for non-EL students were near zero across all subject areas, whereas grades in all subject areas tended to be underpredicted for EL students, with greater underprediction, on average, for EL students who tested with accommodations than for EL students who tested without accommodations.

ACT English scores underpredicted grades of EL students to the greatest extent across subject areas, especially math, underpredicting math grades by one-third to one-half of a grade point. Math grades showed the highest degree of underprediction across ACT subject tests, where ACT English scores underpredicted math grades to the greatest extent (0.34-0.48), and ACT math scores underpredicted math grades to a lesser, but still substantial, extent (0.24-0.27). The only area where underprediction was not found was that ACT math scores did not

underpredict social studies grades, and in fact, may have slightly overpredicted social studies grades of EL students who took the state test with accommodations.

Table 10. Mean Residuals for Actual – Predicted Subject-Specific HSGPA for State Test

	HSGPA	ACT Subject Test			
Testing Group	Subject	English	Math	Reading	Science
Non-EL, No Accommodations	English	0.00	0.00	0.00	0.00
EL, No Accommodations		0.19	0.05	0.15	0.10
EL, State-Allowed Accommodations		0.30	0.05	0.19	0.13
Non-EL, No Accommodations	Math	-0.01	0.00	0.00	0.00
EL, No Accommodations		0.34	0.24	0.27	0.28
EL, State-Allowed Accommodations		0.48	0.27	0.34	0.34
Non-EL, No Accommodations	Social Studies	0.00	0.00	0.00	0.00
EL, No Accommodations		0.15	0.02	0.11	0.07
EL, State-Allowed Accommodations		0.19	-0.03	0.10	0.05
Non-EL, No Accommodations	Science	0.00	0.00	0.00	0.00
EL, No Accommodations		0.22	0.10	0.17	0.15
EL, State-Allowed Accommodations		0.35	0.11	0.23	0.20
Non-EL, No Accommodations	Overall	0.00	0.00	0.00	0.00
EL, No Accommodations		0.23	0.10	0.17	0.15
EL, State-Allowed Accommodations		0.33	0.10	0.22	0.17

Table 11. Residuals for Actual – Predicted Subject-Specific HSGPA for National Test

	HSGPA	ACT Subject Test			
Testing Group	Subject	English	Math	Reading	Science
Non-EL, No Accommodations	English	0.00	0.00	0.00	0.00
EL, No Accommodations		0.19	0.04	0.14	0.08
EL, State-Allowed Accommodations		0.31	0.08	0.21	0.13
Non-EL, No Accommodations	Math	-0.01	0.00	0.00	0.00
EL, No Accommodations		0.34	0.24	0.26	0.25
EL, State-Allowed Accommodations		0.49	0.30	0.36	0.35
Non-EL, No Accommodations	Social Studies	0.00	0.00	0.00	0.00
EL, No Accommodations		0.14	0.01	0.10	0.05
EL, State-Allowed Accommodations		0.20	-0.01	0.13	0.05
Non-EL, No Accommodations	Science	0.00	0.00	0.00	0.00
EL, No Accommodations		0.21	0.09	0.16	0.13
EL, State-Allowed Accommodations		0.36	0.15	0.25	0.21
Non-EL, No Accommodations	Overall	0.00	0.00	0.00	0.00
EL, No Accommodations		0.22	0.10	0.16	0.13
EL, State-Allowed Accommodations		0.34	0.13	0.24	0.19

EL students who took the state test with accommodations and took the national test without accommodations showed a smaller amount of underprediction on the state test, especially performance on the ACT math and reading tests, where underprediction was reduced by about 0.03 grade points across subject areas. These results suggest that the accommodations may have resulted in scores that more accurately reflect student proficiency, especially in math and reading.

Conclusions/Recommendations

ACT recently began providing a limited set of supports for US EL students taking the ACT. To ensure the validity of scores resulting from testing with the supports, we need to provide evidence supporting the three claims below.

1. Allowing accommodations for EL students results in scores that more truly reflect their actual achievement level rather than being confounded with language proficiency.
2. Allowing accommodations for EL students does not give them an unfair advantage (i.e., scores that are higher than their actual achievement level) compared to students who test under regular administration conditions.
3. Allowing accommodations for EL students results in scores that can accurately predict their success in first year college courses as measured by first-year college GPA.

Ultimately, ACT will need to provide evidence about the specific supports that ACT began offering to EL students in the fall of 2017. This report used data available from states using the ACT statewide for accountability and provided accommodations to EL students that were not approved by ACT and thus resulted in scores that were not college reportable. While there were many limitations of the available data, some tentative conclusions can be found.

The results of the score change analyses suggest that testing with state-allowed accommodations slightly increased the ACT scores of EL students on the state test and resulted in lower disparities in performance compared to non-EL students. EL students who took the ACT with accommodations and retested under standard administration conditions within a year after the state test showed smaller gains relative to students who tested twice under standard administration conditions, suggesting that the accommodations may have improved performance on the state test. Correlational analyses showed a much weaker relationship between ACT scores and high school grades for EL students; notably, EL students who tested on a national test date without accommodations had lower correlations between their ACT Composite scores and HSGPA ($r = 0.31$) than when the same students tested during state and district testing and did receive accommodations ($r = 0.36$). Also of note were the much lower correlations between ACT reading scores and other subject area tests for EL students who tested with supports on the state test and without supports on the national test. Regression analyses showed that ACT scores underpredicted the high school grades of EL students, suggesting that their ACT scores did not accurately reflect their actual achievement levels. For EL students who took the ACT with accommodations on the state test and retested under standard administration conditions, underprediction was smaller on the state test. Together, these results suggest that the accommodations allowed these students to more accurately demonstrate their true achievement levels.

However, due to the many limitations of this study, including small sample sizes, and many factors that could not be considered in this study due to incomplete information about students' level of English proficiency, disability status, and types of accommodations used, these results should be considered preliminary. There is not adequate evidence to properly evaluate the

impact of state-allowed accommodations on the scores of EL students. Additional data on EL students who take the ACT test with and without accommodations must be collected, and with sufficient sample sizes to enable us to establish the appropriateness of each type of and combination of ACT-allowed accommodation(s).

In addition, this paper focused on available evidence in comparing the ACT scores of non-EL students and EL students testing with and without accommodations in terms of score changes and relationships with high school grades. As noted earlier, multiple sources of evidence are required to address the issues which emerge concerning the score validity and comparability of ACT scores across these three groups. This paper does *not* address score validity as it relates to college performance (i.e., college grades, GPA), the internal structure of the ACT (e.g., dimensionality, factor structure), or differential item performance. Further evidence should be accumulated to address these and other issues to provide a compelling line of evidence to support the validity of scores for EL students receiving the specific supports approved by ACT.

In terms of predictive evidence, we will need to collect data from a cross-section of post-secondary institutions, including college enrollment, coursework taken, grades, and GPA of EL students who took the ACT under regular and accommodated conditions. This would require either the same students testing both with and without accommodations or equivalent groups of students testing with and without accommodations. Similar studies have been conducted for students with disabilities, and it is hypothesized that the use of appropriate accommodations by EL students may result in scores that are better predictors of first-year college GPA by reducing construct-irrelevant variance due to non-proficiency in English (Cahalan, Mandinach, & Camara,

2002). Conducting such studies will take several years as students initially receiving supports in fall of 2017 will not complete their first year of college until the summer of 2019 or 2020.[3]

Construct continuity and internal structure can be addressed by comparing factor analyses for EL students testing with and without supports. Differential item functioning analyses can be conducted to evaluate whether items are functioning differently for EL and non-EL students testing with and without supports. Omit rates for students taking the ACT on paper and latency analyses for students taking the ACT online can also be compared for non-EL and EL students tested with and without supports to assess the impact of providing extra time. In addition, test precision and reliability can be examined for EL students tested with and without supports.

In conclusion, we have limited evidence that EL students' ACT scores may not reflect their true achievement levels, and providing accommodations or supports may improve their scores and provide a more accurate reflection of their ability in terms of relationships with high school grades. However, much more work is needed to ensure that the ACT-approved supports are allowing EL students to more accurately reflect their actual achievement level, do not provide an unfair advantage, and more accurately predict college success.

[3] This type of study should include both high school juniors and seniors.

References

Abedi, J. (2001). *Assessment and accommodations for English language learners: Issues and recommendations* (CRESST, Policy Brief 4). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometric issues. *Educational Assessment, 8*(3), 231–257.

Abedi, J. (2003). Testing of English language learner students. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J. –I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, vol. 3: Testing and assessment in school psychology and education* (355-368). Washington, DC: American Psychological Association.

Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher*, *33*(1), 4-14.

Abedi, J. (2008a). Classification system for English language learners: Issues and recommendations. *Educational Measurement: Issues and Practice, 27*(3), 17-31.

Abedi, J. (2008b). Measuring students' level of English proficiency: Educational significance and assessment requirements. *Educational Assessment, 13*(2-3), 193-214.

Abedi, J. (2009). English Language Learners with disabilities: Classification, assessment, and accommodation issues. *Journal of Applied Testing Technology, 10*(2), 1-30.

Abedi, J. (2014). The use of computer technology in designing appropriate test accommodations for English language learners, *Applied Measurement in Education, 27*(4), 261-272.

Abedi, J., Courtney, M., & Leon, S. (2003). *Research-supported accommodation for English language learners in NAEP* (CSE Tech. Rep. No. 586). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Courtney, M., Mirocha, J., Leon, S., & Goldberg, J. (2005). *Language accommodations for English language learners in large-scale assessments: Bilingual dictionaries and linguistic modification* (CSE Report 666). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., & Ewers, N. (2013). *Accommodations for English language learners and students with disabilities: A research-based decision algorithm*. Davis, CA: University of California, Davis, Smarter Balanced Assessment Consortium.

Abedi, J., Hofstetter, C., Baker, E., & Lord, C. (2001). *NAEP math performance and test accommodations: Interactions with student language background* (CSE Tech. Rep. No. 536). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research, 74*(1), 1-28.

Abedi, J., & Gandara, P. (2006). Performance of English language learners as a subgroup in large-scale assessment: Interaction of research and policy. *Educational Measurement: Issues and Practice, 25*(4), 36-46.

Abedi, J., Leon, S., & Mirocha, J. (2003). *Impact of student language background on content based performance: Analyses of extant data* (CSE Tech. Rep. No. 603). Los Angeles,

CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14*(3), 219–234. doi: 10.1207/S15324818AME1403_2.

Abedi, J., Lord, C., Boscardin, C.K, & Miyoshi J. (2001). *The Effects of accommodations on the assessment of limited English proficient (LEP) students in the National Assessment of Educational Progress (NAEP)* (CSE Tech. Rep. No. 537). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Lord, C., & Hofstetter, C. (1998). *Impact of selected background variables on students' NAEP math performance* (CSE Tech. Rep. No. 478). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice, 19*(3), 16–26.

Abedi, J., Lord, C., & Plummer, J. R. (1997). *Final report of language background as a variable in NAEP mathematics performance* (CSE Tech. Rep. 429). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Acosta, B, Rivera, C. & Willner, L. S. (2008). *Best practices in state assessment for accommodating English language learners: A Delphi study*. Washington, DC: The

George Washington University Center for Equity and Excellence in Education. Retrieved from: http://files.eric.ed.gov/fulltext/ED539759.pdf

ACT (2017a). ACT policy for accommodations documentation. Iowa City, IA: ACT. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/6368%20ACT%20Policy%20for%20Documentation-Web.pdf.

ACT (2017b). ACT policy for English learner supports documentation. Iowa City, IA: ACT. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/ACTPolicyforEnglishLearnerSupportsDocumentation.pdf

Albus, D., Bielinski, J., Thurlow, M., & Liu, K. (2001). *The effect of a simplified English language dictionary on a reading test* (LEP Projects Report 1). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from https://nceo.info/Resources/publications/OnlinePubs/LEP1.html

Anderson, M., Liu, K., Swierzbin, B., Thurlow, M., & Bielinski, J. (2000). *Bilingual accommodations for limited English proficient students on statewide reading tests: Phase 2* (Minnesota Report No. 31). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from https://nceo.info/Resources/publications/OnlinePubs/archive/AssessmentSeries/MnReport31.html

Andrews, K. M. & Ziomek, R. L. (1998). *Score gains on retesting with the ACT assessment*. Iowa City, IA: ACT. Retrieved from http://www.act.org/content/dam/act/unsecured/documents/ACT_RR98-07.pdf

Bailey, A. L. & Carroll, P. E. (2015). Assessment of English language learners in the era of new academic content standards. *Review of Research in Education, 39*(1), 253-294.

Bridgeman, B., & Cline, F. (2007). *Effects of extra time on performance on new SAT questions* (Research Notes RN-33, ETS RM-07-02). New York, NY: The College Board.

Bridgeman, B., Trapani, C., & Curley, E. (2004). Impact of fewer questions per section on SAT I scores. *Journal of Educational Measurement, 41*(4), 291-310.

Cahalan, C., Mandinach, E. B., & Camara, W. J. (2002). *Predictive validity of SAT I: Reasoning test for test-takers with learning disabilities and extended-time accommodations* (College Board Research Report, 2002-05). New York, NY: The College Board.

Camara, W. J. (2009). Validity evidence in accommodations for English language learners and students with disabilities. *Journal of Applied Testing Technology, 10*(2), 1-23.

Camara, W. J., & Allen, J. (2017). *Does testing date impact student scores on the ACT?* Iowa City, IA: ACT. Retrieved from http://www.act.org/content/dam/act/unsecured/documents/R1643-test-date-impact-on-act-scores-2017-06.pdf.

Camara, W. J., Copeland, T., & Rothschild, B. (1998). *Effects of extended time on the SAT I: Reasoning Test score growth for students with learning disabilities*. College Board Report No. 98-7. New York, NY: The College Board. Retrieved from https://files.eric.ed.gov/fulltext/ED562679.pdf

Cawthon, S., Ho, E., Patel, P., Potvin, D., & Trundt, K. (2009). Multiple constructs and effects of accommodations on accommodated test scores for students with disabilities. *Practical Assessment, Research & Evaluation, 14*(18). Retrieved from http://pareonline.net/pdf/v14n18.pdf

Chiu, C. W. T., & Pearson, P. D. (1999, June). *Synthesizing the effects of test accommodations for special education and limited English proficiency students*. Paper presented at the National Conference on Large Scale Assessment, Snowbird, UT.

Crotts, K. (2013). *Evaluating the validity of accommodations for English learners through evidence based on response processes* (Doctoral dissertation 786). Retrieved from http://scholarworks.umass.edu/open_access_dissertations/786

Ercikan, K., Roth, W. –M., Simon, M., Sandilands, D., & Lyons-Thomas, J. (2014). Inconsistencies in DIF detection for sub-groups in heterogeneous language groups. *Applied Measurement in Education, 27*(4), 273-285.

Francis, D., Rivera, M., Lesaux, N., Kieffer, M., & Rivera, H. (2006). *Practical guidelines for the education of English language learners: Research-based recommendations for the use of accommodations in large-scale assessments*. Portsmouth, NH: RMC Research Corporation, Center on Instruction. Retrieved from http://centeroninstruction.org/files/ELL3-Assessments.pdf

Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C. L., Binkley, E., & Crouch, R. (2000). Using objective data sources to enhance teacher judgments about test accommodations. *Exceptional Children, 67*(1), 67-81.

Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C. L., & Karns, K. M. (2001). Supplementing teacher judgments of mathematics test accommodations with objective data sources. *School Psychology Review, 29(1),* 65-85.

Garcia, G. E. (1991). *Factors influencing the English reading test performance of Spanish-speaking Hispanic children* (Technical Report No. 539). Champaign, IL: University of Illinois at Urbana-Champaign, Center for the Study of Reading.

Hafner, A. (2001, April). *Evaluating the impact of test accommodations on test scores of LEP students & non-LEP students*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Harmston, M., & Crouse, J. (2016). *Multiple testers: What do we know about them?* Iowa City, IA: ACT. Retrieved from http://www.act.org/content/dam/act/unsecured/documents/2016-Tech-Brief-MultipleTesters.pdf.

Helwig, R., & Tindal, G. (2003). An experimental analysis of accommodation decisions on large-scale mathematics tests. *Exceptional Children, 69*(2), 211-225.

Herman, J. L. & Abedi, J. (2004). *Issues in assessing English language learners' opportunity to learn mathematics* (CSE Report 633). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Hill, G. A. (1984). *Learning disabled college students: The assessment of academic aptitude.* (Unpublished Ph.D. dissertation). Texas Tech University, Lubbock, TX.

Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English language learners taking large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research, 79*(3), 1168-1201.

Kieffer, M., Rivera, M., & Francis, D. (2012). *Practical guidelines for the education of English language learners: Research-based recommendations for the use of accommodations in large-scale assessments Book 4.* Portsmouth, NH: RMC Research Corporation, Center on Instruction. Retrieved from http://www.nsbsd.org/cms/lib01/AK01001879/Centricity/Domain/781/Researched-

based%20Recommendations%20For%20the%20Use%20of%20Accommodations%20in%20Large-scale%20Assessments.pdf

Kindler, A. L. (2002). *Survey of the states' limited English proficiency students and available educational programs and services: 2000-2001 summary report*. Washington, DC: National Clearinghouse for English Language Acquisition & Language Instruction Educational Programs.

Kopriva, R. J., Emick, J. E., Hipolito-Delgado, C. P., & Cameron, C. A. (2007). Do proper accommodation assignments make a difference? Examining the impact of improved decision making on scores for English language learners. *Educational Measurement: Issues and Practice, 26*(3), 11-20.

Koran J., & Kopriva, R. J. (2006). *Teacher assignment of test accommodations for English language learners: Does it work?* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA. Retrieved from http://iiassessment.wceruw.org/research/researchDocuments/Teacher%20and%20STELLA%20Assignment%20of%20Accommodations%20Koran%20Kopriva%202006.pdf.

Koran, J., & Kopriva, R. J. (2008). Proper assignment of accommodations to individual students In R. J. Kopriva, *Improving testing for English language learners* (pp. 221-258). New York, NY: Routledge Publishers.

Lane, S., & Levanthal, B. (2015). Psychometric challenges in assessing English language learners and students with disabilities. *Review of Research in Education, 39*(1), 165-214.

Lazarin, M. (2006). *Improving assessment and accountability for English language learners in the No Child Left Behind Act* (Issue Brief. No 16). Washington, DC: National Council of La Raza.

Li, H., & Suen, H. K. (2012). The effects of test accommodations for English language learners: A meta-analysis. *Applied Measurement in Education, 25*(4), 327-346.

Linquanti, R., Cook, H. G., Bailey, A. L., & MacDonald, R. (2016). *Moving toward a more common definition of English learner: Collected guidance for states and multi-state assessment consortia.* Washington, DC: Council of Chief State School Officers.

Lovett, B. J. (2011). Extended time testing accommodations: What does the research say? *Communique: The Newspaper of the National Association of School Psychologists, 39*(8), 1, 14-15. Retrieved from http://web.a.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=3&sid=e0684bfe-d2bc-433f-b5a1-b37ab5f948b5%40sessionmgr4010.

Mandinach, E. B., Bridgeman, B., Cahalan-Laitusis, C., & Trapani, C. (2005). *The impact of extended time on SAT test performance* (College Board Research Report No. 2005-8 / ETS RR-05-20). New York, NY: The College Board.

Mattern, K. D., Patterson, B. F., & Kobrin, J. L. (2012). *The validity of SAT® scores in predicting first-year mathematics and English grades* (College Board Research Report No. 2012-1). New York, NY: The College Board.

Mattern, K. D., Patterson, B. F., Shaw, E. J., Kobrin, J. L., & Barbuti, S. M. (2008). *Differential validity and prediction of the SAT®* (College Board Research Report No. 2008-4). New York, NY: The College Board.

National Center for Education Statistics (2015). *NAEP accommodations increase inclusiveness.* Retrieved from https://nces.ed.gov/nationsreportcard/about/accom_table.aspx.

National Research Council (2000). *Testing English-language learners in U.S. schools: Report and workshop summary. Committee on Educational Excellence and Testing Equity.*

Board on Testing and Assessment, Center for Education. Washington, DC: National Academy Press.

Noble, T., Rosebery, A., Suarez, C., Warren, B., & O'Connor, M. C. (2014). Science assessments and English language learners: Validity evidence based on response processes. *Applied Measurement in Education*, *27*(4), 248-260.

O'Conner, R., Abedi, J., & Tung, S. (2012a). *A descriptive analysis of enrollment and achievement among limited English proficient students in New Jersey* (Issues & Answers, REL 2012-No. 108). Retrieved from http://files.eric.ed.gov/fulltext/ED531432.pdf

O'Conner, R., Abedi, J., & Tung, S. (2012b). *A descriptive analysis of enrollment and achievement among English language learner students in Pennsylvania: Summary* (Issues & Answers, REL 2012-No. 127). Retrieved from http://files.eric.ed.gov/fulltext/ED531429.pdf

Oliveri, M. E., Ercikan, K., & Zumbo, B. D. (2014). Effects of population heterogeneity on accuracy of DIF detection. *Applied Measurement in Education, 27*(4), 286-300.

Partnership for Assessment of Readiness for College and Careers (2017). *PARCC accessibility features and accommodations manual: Guidance for districts and decision-making teams to ensure that PARCC summative assessments produce valid results for all students.* Sixth Edition. Washington, DC: PARCC Assessment Consortia. Retrieved from https://parcc.pearson.com/manuals/

Pennock-Roman, M., & Rivera, C. (2011). Mean effects of test accommodations for ELLs and non-ELLs: A meta-analysis of experimental studies. *Educational Measurement: Issues and Practice 30*(3), 10-28.

Ragosta, M., & Wendler, C. (1992). *Eligibility issues and comparable time limits for disabled and nondisabled SAT examinees* (College Board Report No. 92-5. ETS RR No. 92-35). New York, NY: College Entrance Examination Board. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1992.tb01466.x/pdf

Ramist, L., Lewis, C., & McCamley-Jenkins, L. (1994). *Student group differences in predicting college grades: Sex, language, and ethnic groups* (College Board Research Report No. 93-1. ETS RR No. 94-27). New York, NY: The College Board.

Reschly, D. J. (1996). Identification and assessment of students with disabilities. *The Future of Children, 6*(1), 40-53.

Rivera, C., Acosta, B. D., & Willner, L. S. (2008). *Guide for refining state assessment policies for accommodating English language learners.* Arlington, VA: The George Washington University Center for Equity and Excellence in Education.

Rivera, C., & Collum, E. (2004). *An analysis of state assessment policies addressing the accommodation of English language learners*. NAGB Conference on Increasing the Participation of SD and LEP Students in NAEP: Commissioned Paper Synopsis. The George Washington University Center for Equity and Excellence in Education.

Rivera, C., & Stansfield, C. W. (1998). Leveling the playing field for English language learners: Increasing participation in state and local assessments through accommodations. In R. Brandt (Ed.), *Assessing student learning: New rules, new realities*. Arlington, VA: Educational Research Service and Alliance for Curriculum Reform.

Roohr, K. C., & Sireci, S. G. (2017). Evaluating computer-based test accommodations for English learners. *Educational Assessment, 22*(1), 35-53. Retrieved from http://dx.doi.org/10.1080/10627197.2016.1271704

Sanchez, E., & Buddin, R. (2015). *How accurate are self-reported high should courses, course grades, and grade point average?* Iowa City, IA: ACT. Retrieved from http://www.act.org/content/dam/act/unsecured/documents/WP-2015-03.pdf

Sawyer, R. (2010). *Usefulness of high school average and ACT scores in making college admissions decisions.* Iowa City, IA: ACT.

Schnipke, D. L., & Pashley, P. J. (1997). *Assessing subgroup differences in item response times*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, March 24-28.

Shepard, L., Taylor, G., & Betebenner, D. (1998). *Inclusion of limited-English-proficient students in Rhode Island's Grade 4 Mathematics Performance Assessment* (CSE Tech. Rep. No. 486). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Sireci, S. G., & Faulkner-Bond, M. (2015). Promoting validity in the assessment of English learners. *Review of Research in Education, 39*(1), 215-252.

Sireci, S. G., Li, S., & Scarpati, S. (2003). *The effects of test accommodation on test performance: A review of the literature* (Research Rep. 485). Amherst, MA: Center for Educational Assessment.

Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research, 75*(4), 457-490.

Smarter Balanced Assessment Consortium (2017). *Smarter Balanced Assessment Consortium: Usability, accessibility, and accommodations guidelines*. Davis, CA: Smarter Balanced Assessment Consortium. Retrieved from

https://portal.smarterbalanced.org/library/en/usability-accessibility-and-accommodations-guidelines.pdf

Solano-Flores, G. (2006). Language, dialect, and register: Sociolinguistics and the estimation of measurement error in the testing of English language learners. *Teachers College Record, 108*(11), 2354–2379.

Solano-Flores, G., & Li, M. (2009). Language variation and score variation in the testing of English language learners, native Spanish speakers. *Educational Assessment, 14*(3-4), 180-194.

Solano-Flores, G., Wang, C., Kachchaf, R., Soltero-Gonzalez, L., & Nguyen-Le, K. (2014). Developing testing accommodations for English language learners: Illustrations as visual supports for item accessibility. *Educational Assessment, 19*(4), 267-283.

Steinberg, J., Cline, F., Ling, G., Cook, L., & Tognatta, N. (2009). Examining the validity and fairness of a state standards-based assessment of English-Language Arts for deaf or hard of hearing students. *Journal of Applied Testing Technology, 10*(2), 1-33.

Thurlow, M. L., Larson, E. D., Lazarus, S. S., Shyyan, V. V., & Christensen, L. L. (2017). *Educators' experiences with and attitudes toward accessibility features and accommodations*. Minneapolis, MN: University of Minnesota, Data Informed Accessibility—Making Optimal Needs-based Decisions (DIAMOND).

US Department of Education (2016). *Non-Regulatory Guidance: English Learners and Title III of the Elementary and Secondary Education Act (ESEA), as amended by the Every Student Succeeds Act (ESSA), Appendix A.* Retrieved from https://www2.ed.gov/policy/elsec/leg/essa/essatitleiiiguidenglishlearners92016.pdf.

Weston, T. J. (2003). *NAEP validity studies: The validity of oral accommodation in testing* (Working Paper No. 2003-06). Washington, DC: US Department of Education, National Center for Education Statistics.

Wolf, M. K., Herman, J. L., & Dietel, R. (2010). *Improving the validity of English language learner assessment systems* (CRESST Policy Brief No. 10 - Full Report). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Wolf, M. K., Kao, J., Griffin, N., Herman, J. L., Bachman, P. L., Chang, S. M., & Farnsworth, T. (2008). *Issues in assessing English language learners: English language proficiency measures and accommodations uses. Practice Review (Part 2 of 3)* (CRESST Report 732). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Young, J. W., & King T. C. (2008). *Testing accommodations for English language learners: A review of state and district policies* (College Board Research Report No. 2008-6. ETS Research Report Series, 2008(48)) New York, NY: The College Board.

Young, J. W., Cho, Y., Ling, G., Cline, F., Steinberg, J., & Stone, E. (2008). Validity and fairness of state standards-based assessments for English language learners. *Educational Assessment*, *13*(2-3), 170-192.

Zenisky, A. L., & Baldwin, P. (2006). *Using item response time data in test development and validation: Research with beginning computer users*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA, April 8-10.

Ziomek, R. L., & Andrews, K. M. (1998). *ACT assessment score gains of special-tested students who tested at least twice*. Iowa City, IA: ACT.

Zwick, R., & Schlemer, L. (2004). SAT validity for linguistic minorities at the University of California, Santa Barbara. *Educational Measurement: Issues and Practice, 23*(1), 6–16.

Zwick, R., & Sklar, J. C. (2005). Predicting college grades and degree completion using high school grades and SAT scores: The role of student ethnicity and first language. *American Educational Research Journal, 42*(3), 439-464.