



# Classroom assessment and English Language Learners: Teachers' accommodations implementation on routine math and science tests



Beth Clark-Gareca

Teachers College, Columbia University, 525 W. 120th Street, New York, NY 10027, USA

## HIGHLIGHTS

- Math and science teachers tended not to know the proficiency levels of ELLs in their classes.
- Teachers tended to decrease accommodations implementation as student proficiency increased.
- Accommodations in students' first languages were seldom used during routine tests.
- High-stakes test accommodations were variably implemented during classroom tests.
- Children in Special Education received more accommodations than any other students.

## ARTICLE INFO

### Article history:

Received 26 December 2014

Received in revised form

2 October 2015

Accepted 11 November 2015

Available online 23 December 2015

### Keywords:

Classroom-based assessment

Math and science assessment

Teacher education

Accommodations

ELLs

## ABSTRACT

In a time of unprecedented educational accountability in the U.S., this mixed-methods study was conducted to explore teacher accommodations implementation when assessing ELLs during routine math and science tests. Elementary teachers in ten Pennsylvania school districts ( $n = 213$ ) were surveyed about their testing practices and accommodations use, and interviews were conducted with fourth grade teachers ( $n = 10$ ) about their assessment practices. Findings suggest that teachers implemented and withdrew accommodations based on students' levels of English proficiency, first language accommodations were infrequently implemented in the classroom context, and that ELLs with IEPs received more accommodations than other ELLs. Implications for policy and practice are discussed.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

In the current educational climate in the United States, students in elementary school are taking more tests than ever before. On the heels of No Child Left Behind, educational accountability is on the rise through the Race to the Top program in support of Common Core State Standards in math and language arts intended to make U.S. students more competitive in a newly, global society (Common Core State Standards Initiative, 2013; Maxwell, 2014). Agencies such as the Partnership for Assessment of Readiness for College and Careers (PARCC) and SMARTER Balanced Assessment Consortium have piloted and implemented new Common Core-aligned assessments during the 2014–15 school year. Such initiatives will certainly continue the trend of academic accountability through

assessment for teachers and students into the foreseeable future (PARCC, 2013; Smarter Balanced Assessment Consortium, 2013).

Amidst this surge in testing, English Language Learners (ELLs) inspire additional concern. ELLs are children who have been identified to speak a language other than English at home, and are eligible for specialized language services in school to further their English language proficiency in school. The eventual goal of this assistance is to facilitate students' full participation in English-based school curriculum. In school year 2002–03, ELLs made up 8.7% of the total U.S. public school population; in school year 2012–13, that percentage had risen to 9.2%, and was estimated at 4.4 million students (NCES, 2015). ELLs continue to be one of the fastest growing demographics in schools, a reality that presents new assessment challenges for schools and teachers that have never experienced such linguistic and cultural diversity in their classrooms before (Ferrara & DeMauro, 2006; Young et al., 2008). In many schools, the responsibility for all content assessment, i.e.,

E-mail address: [bkc2004@tc.columbia.edu](mailto:bkc2004@tc.columbia.edu).

testing in subjects such as math, science, social studies, and language arts, falls to classroom teachers who have little background in principles of assessment or second language acquisition (Cizek, 2007). Nonetheless, these teachers are charged with the task of evaluating the content mastery of all students, including ELLs and non-ELLs alike.

Formal classroom assessment, in the form of quizzes and tests, is a primary means by which ELL academic achievement is evaluated. Defined as “the collection, evaluation, and use of information to help teachers make better decisions” (McMillan, 2004, p. 8), classroom assessment is an important source of information by which students' scholastic performance is measured. These scores affects subsequent, consequential school and classroom-level decisions like assigning grades (Gottlieb, 2006; Willingham, Pollack, & Lewis, 2002), making course placement decisions (LaCelle-Peterson & Rivera, 1994; Oakes, 2005), and identifying students with learning difficulties who may need additional intervention (McMillan, 2004).

Up to this point, the literature related to classroom assessment for ELLs has focused heavily on assessments of English language proficiency. Within the language testing literature, the majority of classroom-based work has related to English proficiency tests (Brindley, 1998; Davison, 2004; Davison & Leung, 2009; Llosa, 2012; 2008; 2007; 2005; McNamara, 2001; Rea-Dickins & Gardner, 2000; Rea-Dickins, 2001; 2004), which has left questions about ELL school assessment in content areas like math and science largely unanswered. In response to this focus on testing practices, a scholarly interest in test accommodations for ELLs has developed; however, its focus up to this point has been primarily directed toward high-stakes tests (Abedi & Lord, 2001; Abedi, Lord, Hofstetter, & Baker, 2000; Kieffer, Lesaux, Rivera, & Francis, 2009; Li & Suen, 2012; Rivera, Stansfield, Scialdone, & Sharkey, 2000; Stansfield & Rivera, 2001; Wolf, Kim, & Kao, 2012) rather than classroom contexts.

Despite the strong influence of high-stakes tests on elementary classroom environments, to date, studies at the nexus of content classroom assessment of ELLs have been relatively sparse. In today's schools, pull-out instruction in the form of language arts replacement is a very common means by which English as a Second Language (ESL) services are provided (Diaz-Rico & Weed, 2002). New inclusionary efforts, such as push-in structures where ESL instruction is provided through a co-teaching model, support ELLs spending more of their time in a general classroom. These developments likely have positive aspects for ELLs in that they can gain access and exposure to mainstream content curriculum, but also may result in limited participation in English-based, unaccommodated tasks designed for proficient English speakers (Diaz-Rico & Weed, 2009; Harper & deJong, 2004; Menken, 2006). The studies that have investigated how ELLs fare in mainstream classrooms have focused in large part on how teachers provide, or don't provide, access to *instruction* rather than assessment (Cho & Reich, 2008; Harklau, 1994; Reeves, 2004). Additionally, with classroom tests as routine occurrences in math and science classrooms, the content classroom becomes a compelling context in which to conduct research about assessment practices with ELLs.

The present study addresses the area of inquiry of accommodations implementation for ELLs on classroom math and science tests. To use Acosta, Rivera, and Shafer-Willner's definition, test accommodations for ELLs refer to “changes to testing procedures, testing materials, or the testing situation in order to allow students' meaningful participation in the assessment” (Acosta, Rivera, & Shafer Willner, 2008, p.vii). Providing ELL test-takers with standard accommodations, such as bilingual dictionaries or additional time, has been suggested to be a beneficial practice when assessing ELLs on high-stakes measures. Implementing accommodations

may lessen the linguistic complexity of a test and allow ELLs to better demonstrate what they know (Abedi & Lord, 2001; Abedi et al., 2000). Though the overall effectiveness of accommodations is a matter of debate (Kieffer et al., 2009; Li & Suen, 2012; Pennock-Roman & Rivera, 2011), teachers are required to implement regulated, standardized accommodations on high-stakes tests as described by state guidelines.

Interestingly, many state departments of education in the United States (e.g., Pennsylvania, Texas, North Carolina, and Florida) call upon accommodations implementation in the *classroom* as a precursor for high-stakes accommodations practice during standardized tests. Taking Pennsylvania as an example, the Pennsylvania guidelines for administration of the Common Core-aligned tests (Pennsylvania System of School Assessment (PSSA) and Keystone Exams) clearly specify the importance of classroom assessment accommodations as a model for accommodations use on high-stakes tests, “Current accommodations used in day-to-day instruction and assessment are appropriate. New accommodations unfamiliar to students should not be introduced to students for the first time when they are taking the PSSA or Keystone Exams” (PDE, 2014, p. 5). To date, however, there has been limited empirical examination of exactly how, or indeed *if*, routine content tests are typically modified at all. As Cizek (2007) has provocatively stated.

The state of research on what constitutes an appropriate accommodation for a given pupil on a large scale assessment is still in its comparative infancy— which would mean by extension that the state of affairs in classroom assessment accommodations is essentially embryonic. (p. 112).

The aim of this study was to contribute to the literature relating to classroom-based assessment and accommodations implementation for ELLs by elementary content teachers. The following research question and sub-questions guided this study:

- 1) What are the reported accommodations implemented by elementary teachers for English Language Learners on classroom math and science tests?
  - a) Do teachers report to change accommodations implementation for ELLs at different proficiency levels?
  - b) Do teachers report to implement accommodations differently for ELLs with special needs than for other ELLs?

## 2. Theoretical framework

Within the field of language assessment, validity theory was used as a frame of reference for this study. For tests to be said to valid forms of measurement, score interpretations and the subsequent decisions made based on those scores must be defensible through a process of logical argument (Bachman & Palmer, 2010). In conjunction with validity concerns, reliability or consistency in measurement of a defined construct is a necessary component of tests because reliable assessment tools allow test results to be interpreted meaningfully (Bachman & Palmer, 1996; 2010).

A relevant component of validity theory is Bachman and Palmer's notion of test usefulness, which suggests that tests are valid if they are “developed with a specific purpose, a particular group of test takers, and a specific language domain ... in mind” (Bachman & Palmer, 1996, p. 18). Because of the fact that most standardized and routine classroom tests are designed with *native* English speakers in mind (Rivera & Collum, 2006), finding ways to appropriately and accurately assess the academic knowledge of ELLs, without changing the intended test construct, presents a significant challenge for classroom teachers. Construct-irrelevant variance is a prevailing concern when testing ELLs, and refers to high correlations between students' content performance and their

**Table 1**  
Participant school district demographics.

School District	Total Student Population	School context	ELL population in school	Total free/Reduced lunch	Reached adequate Yearly Progress	Response rate
1	10,474	Suburb, small	9.30%	63.2%	No	86%
2	15,306	City, small	8.07%	47.6%	No	14%
3	6725	City, small	5.00%	65.7%	No	43%
4	1575	Town, fringe	7.42%	53.6%	Yes	28%
5	4352	Suburb, large	5.74%	41.6%	Yes	15%
6	4431	City, small	13.56%	74.3%	No	12%
7	8391	City, small	7.47%	86.1%	No	
8	12,504	City, mid-size	6.80%	73.1%	No	47%
9	6017	Suburb, large	5.30%	41.3%	Yes	15%
10	11,591	City, small	17.51%	80.8%	No	11%

English proficiency levels resulting in the risk of content tests becoming defacto language proficiency tests (Abedi, 2004; Koretz & Hamilton, 2006; Messick, 1989; Rivera & Collum, 2006). Test accommodations, or changes in tests, testing environments, and scoring criteria, are introduced with the intention of mitigating the effects of construct-irrelevant variance by removing some of the linguistic barriers that ELLs face, and allowing them to be tested fairly with measures that assess the same constructs as their non-ELL classmates.

Some resistance to accommodations has resulted in the current climate of accountability. Teachers are often conflicted between the impetus they feel to administer tests uniformly as standardization requires (Koretz & Hamilton, 2006), or to depart from that standardization and allow ELLs and other students with special needs to participate in high-stakes assessment with accommodations. Accommodations use in low-stakes tests is typically individualized in selection, implementation, and withdrawal, and is often determined at the teacher's discretion (Cizek, 2007). Direct accommodations involving linguistic changes to tests have been found to be more effective in lessening the linguistic burden of tests for ELLs than non-direct accommodations, e.g., changes to the process of administration (Pitoniak et al., 2009; Rivera & Collum, 2006). If accommodations during testing do not interfere with the tested construct, a test can be considered valid as a measure of ELL content knowledge, and scores can be interpreted as meaningful indicators of academic achievement.

### 3. Method

#### 3.1. Context and participants

Pennsylvanian public schools were selected for this study because of the growth the state has experienced in ELL populations in recent years. As of 2012–13 school year, Pennsylvania educated 52,100 ELLs, or 3% of all students in public school, representing an increase of 69% from the decade before (Migration Policy Institute, 2015). Points of comparison are neighbor states New Jersey, Ohio, and New York, with ELL populations at 4.5%, 2.6%, and 8.8% respectively (Migration Policy Institute, 2015). Many of these new immigrants to Pennsylvania are reported to be settling in smaller city, suburban and rural communities (NCES, 2015), where schools have not previously had the need for robust ESL programs.

For the survey portion of the study, participants were elementary teachers ( $n = 213$ ) who were teaching kindergarten through sixth grades in 10 Pennsylvanian school districts. Eighteen of Pennsylvania's 104 school districts met the study's inclusion criteria in that they were located in small city, suburban or town districts per Census determined metro-centric locale categories and had ELL populations that comprised at least 5% of the district's

total student body, which was greater than or equal to 100 ELLs within the district. This stipulation was set to include districts who had sufficient numbers of ELLs so as to increase the likelihood that the teachers would know something about ELLs in their schools. Table 1 details the 10 included school districts.

As Table 1 illustrates, all participant school districts had between 5.3% and 17.5% total ELL populations. Districts had moderate to high rates of free and reduced lunch (41–86%), and seven of the 10 surveyed districts (70%) had not reached Adequate Yearly Progress benchmarks, an annual measure of school performance for the year of data collection. Survey response rates varied widely, from 11% in District 10–86% in District 1. In District 7, for reasons of privacy, school administrators did not release contact lists to the researcher, but rather, administered the survey internally through a school-based listserv. At the conclusion of data collection, the response rate was not calculable due to uncertainty as to the participant reach of the listserv.

For the qualitative portion of the study, School District 1 and School District 2 agreed to serve as focal sites for data collection. Considering that a number of quantitative accommodations studies have focused particularly on fourth graders' performance (Abedi, Courtney, & Leon, 2003; Rivera & Stansfield, 2004), the qualitative perspective of this study serves to expand on prior work with this age group. At the time of data collection, fourth graders in Pennsylvania were required to take part in high-stakes math and science tests regardless of their English proficiency or number of years of residence in the U.S. (U.S. Department of Education, 2004)<sup>1</sup>; therefore, they were usually fully included in math and science classroom assessment as a precursor to their participation in state tests.

In School Districts 1 and 2, with guidance from school administrators and ESL coordinators, seven schools were selected, and all fourth grade teachers in these schools who met the following inclusion criteria were invited to participate: 1) They taught and tested math or science, 2) They had two or more ELLs in their classrooms, and 3) They were willing to be in close communication with the researcher for coordination of classroom data collection. This process resulted in the participation of 10 focal fourth grade teachers in classroom observations and one-on-one interviews. Pseudonyms were assigned and are implemented throughout.

The 231 surveyed respondents met the study's stated inclusion criteria in that they 1) were kindergarten–sixth grade classroom teachers, 2) had ELLs in their classrooms during data collection, 3) assessed ELLs along with other students in math and/or science,

<sup>1</sup> Most students are also required to take the PSSA standardized test in reading, though depending on their length of residency in the U.S., ELLs may be exempt from the reading section (PDE, 2014).

and 4) completed the survey in its entirety.

Table 2 displays the teacher background information for surveyed and focal participants:

As Table 2 indicates, surveyed teachers were relatively evenly dispersed across elementary grade levels, with most teachers reporting to teach fourth grade (18%), followed by first (15%) and second grades (15%). The majority of the teachers were certified in elementary education (96%), with 24% of teachers also holding early childhood certification, and 10% holding ESOL certification. Seventy percent of respondents reported that they held a Masters degree or higher. Fifty-three percent of teachers reported to have fewer than ten years experience in the classroom.

The 10 fourth grade focal teachers included in the study, six women and four men, were all elementary certified, with one focal teacher being also dually certified in reading. Ninety percent of focal teachers held a Master's degree or higher. Focal teachers' years of classroom experience ranged from 6 to 30 years. Of these 10 teachers, three provided science instruction and assessment to ELLs, five provided math instruction and assessment, and two provided both math and science instruction and assessment. They are referred to by pseudonyms throughout the study.

Surveyed and focal teachers were asked about their training in assessment and multicultural education, as well as about their foreign language proficiency. Teachers were asked to gauge their foreign language proficiency based on their assessment of whether their foreign language skills allowed them to adequately communicate in that language. Table 3 displays this information for survey respondents and focal teachers.

Most surveyed participants (59%) reported having taken at least one course in assessment with only 9% of surveyed teachers claiming to have had no formal assessment training. Though two-thirds of teachers (66%) reported taking some coursework related to multicultural education during their training, another 20% of teachers reported never having taken a course in this area. Ninety-six percent of surveyed teachers reported only having English language proficiency: of those teachers, 59% self-identified as having no foreign language skills, and another 37% claimed that despite studying a foreign language, they did not possess fluent

**Table 2**  
Survey and focal teachers' background information.

Background	Survey teachers n = 213	Focal teachers n = 10
Grade Taught		
K	23 (11%)	
1	32 (15%)	
2	32 (15%)	
3	27 (13%)	
4	40 (18%)	10 (100%)
5	28 (13%)	
6	31 (14%)	
Area of Certification		
Elementary educ.	205 (96%)	10 (100%)
Early childhood	52 (24%)	
Reading	25 (12%)	1 (10%)
Spec education	23 (11%)	
ESOL	22 (10%)	
Other	32 (15%)	
Degree Held		
Bachelors	64 (30%)	1 (10%)
Masters	89 (42%)	6 (60%)
Masters +	60 (28%)	3 (30%)
Years of Experience		
1–5	54 (25%)	
6–10	60 (28%)	3 (30%)
11–15	37 (17%)	3 (30%)
16–20	22 (10%)	2 (20%)
Greater than 20	40 (19%)	2 (20%)

**Table 3**  
Survey and focal teachers' education.

Teacher education	Survey teachers n = 213	Focal teachers n = 10
Courses in Assessment		
0	18 (9%)	
1–2	78 (37%)	5 (50%)
3 or more	47 (22%)	2 (20%)
Don't remember	70 (33%)	3 (30%)
Courses in Multi-Cultural Educ.		
0	42 (20%)	1 (10%)
1–2	113 (53%)	8 (80%)
3 or more	27 (13%)	
Don't remember	31 (15%)	1 (10%)
Foreign Language Proficiency		
Monolingual	125 (59%)	5 (50%)
Studied foreign language	79 (37%)	4 (40%)
Bilingual or multilingual	8 (4%)	1 (10%)
Other	1 (1%)	

communication skills.

In the focal group, five teachers (50%) had taken one or two courses in assessment over the course of their careers, and eight teachers (80%) had taken one or two courses in multicultural education. Similar to the surveyed participants, nine teachers (90%) in the focal group reported to be monolingual or to have minimal foreign language skills.

#### 4. Data collection and analysis

For the quantitative portion of the study, an online adaptive survey consisting of 51 items was developed, piloted and conducted with participant teachers to learn more about their assessment practices with ELLs in the areas of math and science. The term *ESL student* was used consistently within the survey because it was the designation used by elementary school teachers to refer to ELLs within public systems in Pennsylvania. The online survey was conducted through surveymonkey and distributed to teachers through school email lists provided by district staff. Themes investigated in Cho and Reich's (2008) study such as grading and teacher accommodation of ELLs were used as a starting point to capture classroom teachers' opinions about classroom math and science assessment. The survey was designed by the researcher as part of a larger dissertation study in collaboration with university advisors and mentors. Survey sections included the following: 1) screening for inclusion, 2) specific accommodation implementation with ELLs at different levels, 3) and teacher educational background related to professional development and coursework. Section two, which investigated accommodations implementation, was the most detailed in that it asked teachers to document their practices with ELLs. Teachers reported on five survey subscales as to whether they had ELLs at four specific proficiency levels as determined by the World-class Instructional Design and Assessment (WIDA) Access for ELLs, the annual language proficiency test given in Pennsylvania to monitor progress toward English language acquisition (WIDA, 2014). The four proficiency levels were the following: Beginner (Level 1–2, WIDA levels *Entering and Beginning*), Intermediate (Level 3–4, WIDA levels *Developing and Expanding*), Advanced (Level 5, WIDA level *Bridging*), and Monitor (Level 6, WIDA level *Reaching*). These categories were also a school-based indicator to determine the provision and allocation of ESL instructional service hours for ELLs in Pennsylvania, with Beginners receiving the greatest amount of specialized instruction. All students who were identified as ELL were entitled to accommodations on high-stakes, standardized tests. The fifth category, Dually identified students, referred to ELLs who also had been identified as

qualifying for special education services, and therefore had legally-protected accommodations plans outlined in their Individualized Education Programs (IEPs).

Survey responses were captured in several different ways: yes/no questions, three-point Likert scale items, with options of *Always*, *Sometimes*, or *Never*, or five-point Likert scale items with options of *Always*, *Often*, *Sometimes*, *Rarely*, and *Never*. Yes/no questions, such as, *On math and/or science classroom tests (not PSSA) do you make any changes to tests or test administrations to help BEGINNER level ESL students take them more easily?*, and three-point Likert scale questions, such as, *How often are BEGINNER Level ESL students (Levels 1–2) required to take math or science tests administered in your classroom?*, were used as indicators for inclusion into the next section, and triggered skip-logic functions of the survey. A five-point scale was used consistently for the five subscales to understand the prevalence of teacher accommodations implementation. A five-point scale was selected for these subscales because it was commonly used for in-house school surveys and had the greatest degree of familiarity for the teacher participants.

The adaptive nature of the survey resulted in variation in the number of responses to each question. For example, if a teacher reported that s/he did not have students at a beginner proficiency level in his/her class, the survey would move ahead to the next relevant section; if a teacher had multiple proficiency levels represented in his/her classes, s/he was prompted to respond to questions for ELLs at each level. Participant numbers were teachers who reported assessing Beginner ELLs in their classrooms ( $n = 161$ ), teachers who assessed Intermediate ELLs ( $n = 155$ ), Advanced ELLs ( $n = 89$ ), Monitor ELLs ( $n = 87$ ), and Dually identified ELLs ( $n = 77$ ).

Internal consistency measures were calculated for the accommodations subscales. All subscales consisted of six items. Cronbach's alpha was calculated for the Beginner subscale ( $\alpha = .670$ ), the Intermediate subscale ( $\alpha = .670$ ), the Advanced subscale ( $\alpha = .746$ ), the Monitor subscale ( $\alpha = .822$ ), and the Dually Identified scale ( $\alpha = .740$ ). All of these values were found to be within acceptable or good ranges, and suggested there was internal consistency between survey items.

Before launching the survey, it was pilot-tested in seven Pennsylvania school districts that had fewer ELLs (2.5%–4.9% ELL within the total school population) than the schools used for the larger study. In the pilot phase, constructed response options were incorporated within the survey, which requested feedback from participant teachers on the clarity of items. Based on the suggestions offered by the participants, the survey was revised to reduce ambiguity in terms, e.g., ELL was changed to ESL student, and clearer definitions of what constituted specific accommodations were included. The pilot also served to determine if the skip-logic was functioning as intended.

For the qualitative portion of the study, semi-structured individual interviews were conducted with the 10 focal fourth grade teachers to learn more about their assessment practices, their knowledge and understanding of tests and accommodations, and the principles that guided their decisions for ELLs. These individual interviews lasted approximately 45 min, and were conducted in the style of conversational partnership (Rubin & Rubin, 2005). The interview protocol consisted of 15 questions, which were divided into four sections, 1) General Assessment of ELLs, 2) Assessment Practices in a Specific Context, which related to the classroom test observations that had been conducted, 3) Accommodations Implementation, 4) and Grading. Data were subsequently coded based on a framework developed by Lee (2004). In her study, Lee (2004) looked at the degree to which professional development in school could help shape elementary teacher beliefs, and coded beliefs relating to teachers' beliefs about children's abilities to learn, academic content, the role of language and culture in instruction,

teachers' self efficacy, and modifications in teaching practices. For the purposes of this study, similar to Lee (2004), teacher beliefs were coded according to the following four categories: teachers' beliefs about students, the academic content of math and science, the role of language acquisition in assessment, and beliefs about assessments, accommodations, and grading practices. Before launching the study, the interview protocol was piloted with two non-participant teachers, and was subsequently compressed to fit into a teacher's planning period. All interviews were digitally recorded and transcribed shortly thereafter for ease of coding and analysis (Seidman, 2006).

## 5. Findings

### 5.1. Teacher report of assessment practices

Surveyed teachers were asked to report their testing procedures related to ELLs. When asked about their students' language proficiency, the majority of surveyed math and science teachers (65%) reported that they were not certain of the WIDA scores or proficiency levels of the ESL students in their classrooms.

In an effort to learn if test exemption was a common practice during content assessments, surveyed teachers were asked the following question for all five categories of ELLs in their classrooms: *How often are (Beginner, Intermediate, Advanced, Monitor, Dually identified) ESL students required to take math or science tests administered in your classroom?* These data are found in Fig. 1.

As Fig. 1 indicates, most surveyed teachers reported always giving math and science tests to ELLs in regular classrooms at all proficiency levels. Teachers reported that they tested students at the Beginner level less frequently than students at Intermediate, Advanced, and Monitor levels. In fact, more than a quarter (27%) of teachers with Beginner level ELLs reported that Beginners were not always required to take math and science tests in their classrooms, a finding which differs from the numbers of teachers who occasionally exempted Intermediate (8%), Advanced (7%), and Monitor (3%) students from their tests. Teachers with Dually identified students also reported assessing these students less frequently in mainstream classrooms, with 23% of teachers reporting to always or occasionally exempt these students from tests.

Two focal teachers reported occasionally exempting Beginner level ELLs from math or science assessment, and supplied reasons that students didn't have the language level to participate in the test, and that they could not be fairly assessed because they had missed the instruction leading up to the test because they were in ESL class.

Besides exempting students from content tests, surveyed

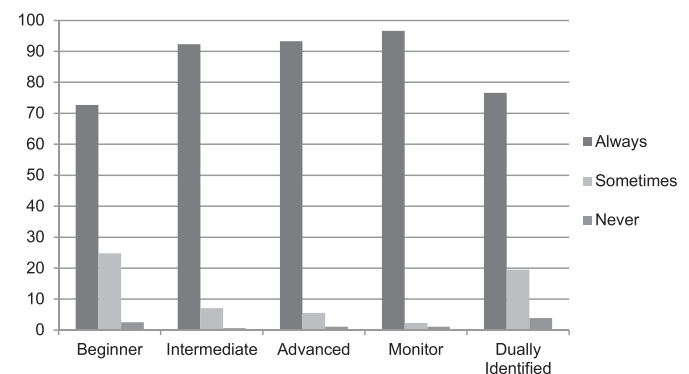


Fig. 1. How often are \_\_\_\_\_ ESL students required to take math or science tests administered in your classroom?

teachers were also asked if they implemented accommodations for ELLs on regular math or science tests. The survey question posed was: *On math and/or science classroom tests, do you make any changes to tests or test administrations to help ESL students take them more easily?* Fig. 2 displays these results for ELLs at different proficiency levels and for Dually identified students.

As seen in Fig. 2, many teachers reported changing tests for ELLs, and the percentages of teachers who implemented accommodations for ELLs decreased as ELL proficiency levels increased. Eighty-three percent of teachers who had Beginner level ELLs claimed to accommodate for them in assessment situations, compared to 68% of teachers with Intermediate, 49% of teachers with Advanced, and 38% of teachers with Monitor level ELLs in their classes. Despite implementing accommodations less frequently, one-third of teachers with Monitor level students (38%) still reported making some changes for them on content assessments. Almost all teachers with Dually identified students (92%) reported implementing accommodations for them during mainstream math and/or science tests.

Similar to the survey results, all 10 focal teachers reported that they were willing to make changes, i.e., accommodations, to help ELLs access the language of content assessment. All focal teachers reported giving more accommodations to ELLs at lower proficiency levels than higher levels because they felt these students tended to struggle more in math and science. Mrs. Butterman, a science teacher, reported a belief that teachers should adjust their assessment expectations to students' proficiency levels, "Well, it depends how they're classified in terms of their ESOL level. You know, certainly a Beginner is going to have a completely different set of issues than someone who's Advanced."

Mr. Baker, a math teacher, reported being less inclined to provide language accommodations for ELLs specifically but did modify assessments as necessary on an individual level:

I wouldn't say I accommodate specifically based on language. I more try to accommodate on strengths and weaknesses of skills ... I don't take it from an English angle, but, if there's a little kid who really just can't understand me at all, I might give him or her just some very simple like 3 + 5 papers for a while. Just to make them comfortable.

Though their rationales for implementation differed, all of the focal teachers reported accommodations to be a regular part of their classroom assessment practices.

### 5.2. Teacher report of specific accommodations implementation

Surveyed teachers were asked about their implementation of a variety of specific accommodations for ELLs on math or science

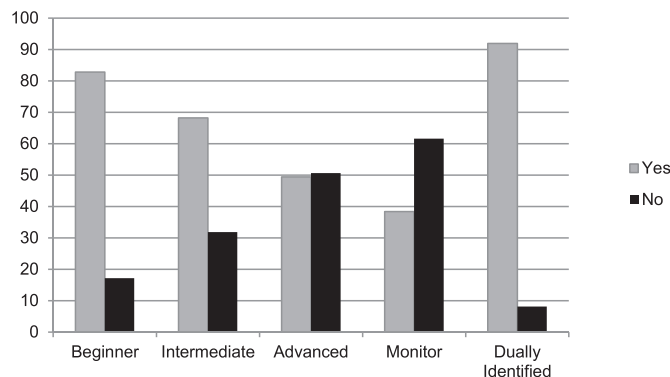


Fig. 2. On math and/or science classroom tests, do you make any changes to tests or test administrations to help ESL students take them more easily?

tests. These accommodations included additional time, teacher assistance, translator/interpreter use, English–English dictionary use, bilingual dictionary use, and bilingual tests. These results are reported by individual accommodation in the next section.

#### 5.2.1. Additional time

The accommodation of additional time was defined within the survey as the practice of giving students additional time to finish the test. The survey question was posed as follows: *How often do you provide additional time to ESL students on math or science tests?* Fig. 3 summarizes the surveyed teachers' report of the frequency of extra time implementation.

As Fig. 3 indicates, very high numbers of teachers reported always offering ELLs additional time on tests. This was the case for ELLs at all proficiency levels and for students who were Dually identified. The frequency with which teachers offered extra time decreased as English proficiency developed, with the *always* response moving steadily downward from 81% for Beginners, 78% for Intermediates, 61% for Advanced, and 58% at the Monitor level. Implementation of additional time for students who were Dually identified was the highest level of any group, with 90% of teachers reporting always offering additional time to these students.

Among the focal teachers, all 10 reported giving extra time on tests to ELLs if they needed it to complete assessment tasks. Mr. Rockne, a math teacher, discussed implementation of additional time in his classroom, "If they need additional time, I give it to them. I have no problem doing that." Extra time was found to be a very common assessment practice across all focal classrooms, schools, and districts.

#### 5.2.2. One-on-one teacher assistance

Teacher assistance was defined within the survey as assistance provided during the test by the ESL teacher or other specialized personnel. The question posed to participants was: *How often do you provide one-on-one teacher assistance to ESL students on math or science tests?* Fig. 4 summarizes the surveyed teachers' report of implementation of teacher assistance with ELLs.

Approximately 30% of surveyed teachers reported always giving one-on-one assistance to Beginner and Intermediate level ELLs. Advanced and Monitor level ELLs were offered teacher assistance on assessments less frequently, with 22% and 29% of these teachers reporting to never assist students at this level. Fifty-nine percent of teachers reported to always provide this accommodation to Dually identified students.

One focal science teacher, Mrs. Lowe, discussed her practice of offering assistance to ELLs as an opportunity to clarify confusion, "For students who don't understand the language as well as others

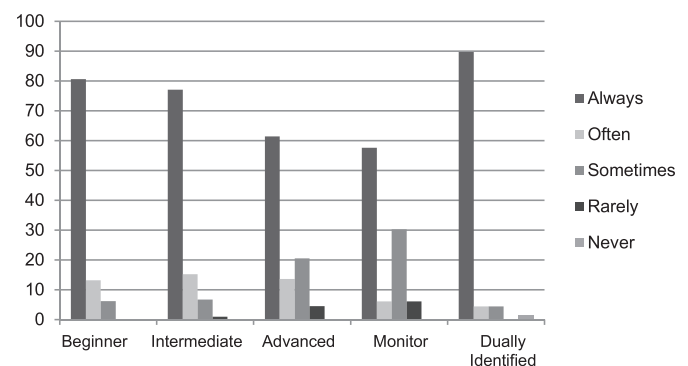


Fig. 3. How often do you provide additional time to ESL students on math or science tests?

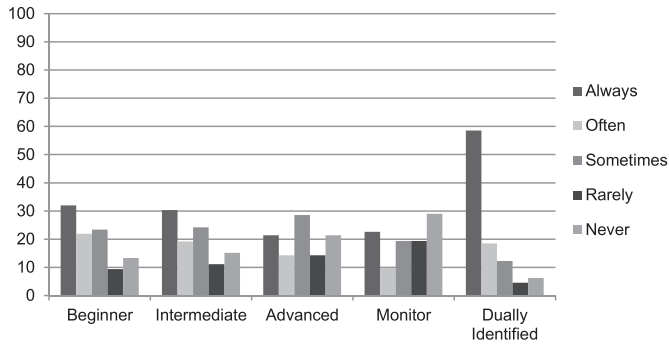


Fig. 4. How often do you provide one-on-one teacher assistance to ESL students on math or science tests?

- then it's easier for them if they're one-on-one with someone who's speaking with them and they'll feel more comfortable asking questions if they don't understand something."

One reported constraint of this accommodation related to the number of available adults to provide one-on-one support. Five focal teachers indicated in their interviews that, at times, accommodations for ELLs were administered by special education professionals in their classrooms. Mr. Baker explained the way the special education support staff worked in his room, "You know they're very flexible as far as whom they'll help. You know, they don't say, okay I'm just going to help my IEP kids. You know, they just help whoever needs help."

5.2.3. Translator/interpreter use in the classroom

Surveyed teachers reported infrequently using a translator or interpreter to translate part or all of a test. The question asked of participants was: *How often do you provide a translator or interpreter for ESL students on math or science tests?* Fig. 5 represents teacher report of translator or interpreter implementation in their classrooms during math or science tests.

At all levels, most surveyed teachers reported rarely or never implementing this accommodation. Rates of implementation were reported as being somewhat higher at the Beginner level than at other levels, but this accommodation was still reported to be uncommon across all proficiencies.

Like survey teachers, focal teachers' opinions about translation seemed to change along students' proficiency levels, in that Beginners were more likely to need, and subsequently receive, such an accommodation. Two of three focal teachers who had newly arrived Beginner level ELLs in their classrooms expressed their frustration and their need for better access to translators for ELLs on assessments. Mr. Roswell, a math teacher, expressed his empathy

for his newly arrived students and his inability to meet their linguistic needs:

[It's] tough too because they're young kids. They don't speak English. They're in a new country, and we just throw them in a classroom. I mean it has to be extremely difficult for them, like you can imagine what they must go through. I just wish there were other things I could do to help them. I oftentimes feel I'm not doing them a full service because I don't speak Spanish; I rely on other people to help me.

Mrs. Butterman and Mr. Smith expressed that translation or interpretation services were not necessary beyond the Beginner level, because students with higher proficiency levels knew enough English to be able to take tests without that support.

5.2.4. Dictionary use

Surveyed teachers were asked about English–English and bilingual dictionary provision as an accommodation on tests. An English–English dictionary was defined within the survey as having words and definitions in English only. A bilingual dictionary was defined as having words and definitions that were translated to the student's native language. Two survey questions were asked about dictionary use: *How often do you allow English–English dictionary use for ESL students in math or science?* and *How often do you allow bilingual dictionary use for ESL students in math or science?* These results are presented in Figs. 6 and 7.

The use of any kind of dictionary by ELLs during tests was reported to be infrequent across proficiency levels. More than 50% of teachers reported never to implement English–English or bilingual dictionaries for their students during math and science tests. Unlike other accommodations investigated in this study, the reported implementation of both types of dictionaries increased as student proficiency increased; that is, the accommodation of dictionary support was offered more frequently to Monitor level ELLs than to any other proficiency level.

In the focal data, three teachers reported allowing ELLs to use English–English dictionaries on tests, and one teacher reported the use of bilingual dictionaries in class. Mrs. Butterman did not provide dictionaries to her students, saying that their use required many prerequisite skills, "If they don't know how to spell the word, they can't find it in the dictionary." In addition to troubles with spelling, she believed that ELLs in fourth grade would not likely have the mastery of alphabetical order to make dictionary use beneficial.

Similarly, Mrs. Pally, a science teacher, thought all dictionaries were too difficult for students at such a young age to use effectively, "But, they have a little hard time with dictionaries. Very hard. They

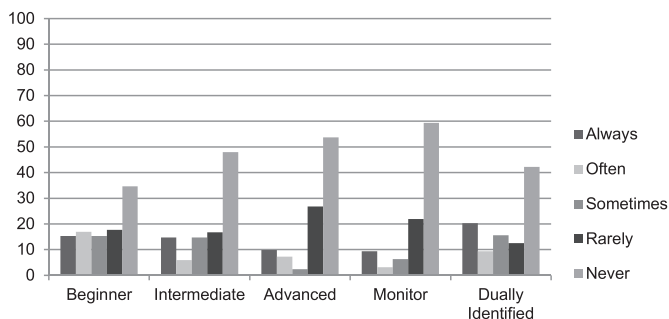


Fig. 5. How often do you provide a translator or interpreter for ESL students on math or science tests?

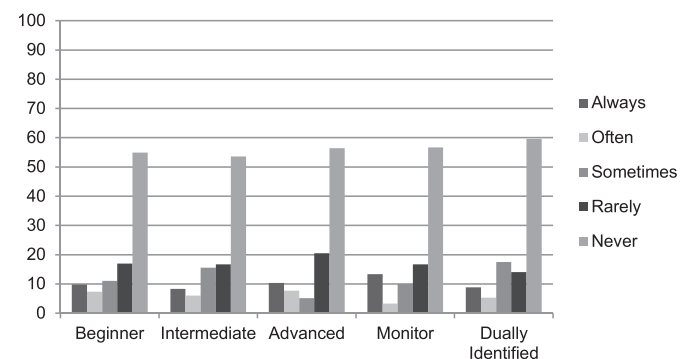


Fig. 6. How often do you allow English–English dictionary use for ESL students on math or science tests?

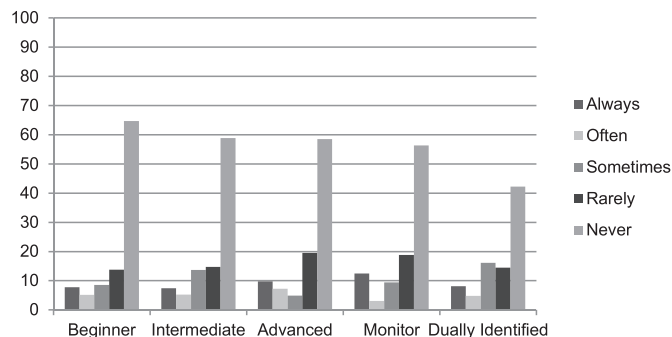


Fig. 7. How often do you allow bilingual dictionary use for ESL students on math or science tests?

have to learn how to look them up and ... Maybe in middle school. Maybe.”

Bilingual dictionaries were discussed in focal interviews with all 10 teachers, and were reported to be in use in only one classroom. A common reason for not using bilingual dictionaries, reported by Mrs. Mathers, Mrs. Butterman, and Mrs. Simpson, was that they were not available in their schools. Mrs. Mathers, a math and science teacher, spoke to this lack of resources, “Actually, I don’t have any bilingual dictionaries. There really should be one.” Two other teachers discussed how bilingual dictionaries were available in their schools for teachers to use, but not for students.

5.2.5. Bilingual tests

Bilingual tests, defined within the survey as tests with questions written in both English and the language of the test-taker or only in the language of the test-taker, were reported to be rarely offered to ELLs. The survey question was posed as follows: *How often do you provide bilingual tests for ESL students in math or science?* Fig. 8 summarizes their use as reported by surveyed teachers.

Teachers of ELLs at all levels reported infrequently providing bilingual tests in math or science. Irrespective of proficiency level or dual identification, in the majority of surveyed teachers’ classrooms, bilingual tests were rarely or never offered.

Focal teachers were asked about their use of bilingual tests, and similar to surveyed teachers, none reported using bilingual tests to any extent in their classrooms. Reasons for not using bilingual tests included a lack of bilingual resources to develop tests in ELLs’ native languages and a belief that translated tests were generally not necessary. In some cases, the schools’ ESL teachers were mentioned as a potential bilingual resource during assessments, but no focal content teacher reported having collaborated to develop a bilingual test with an ESL professional.

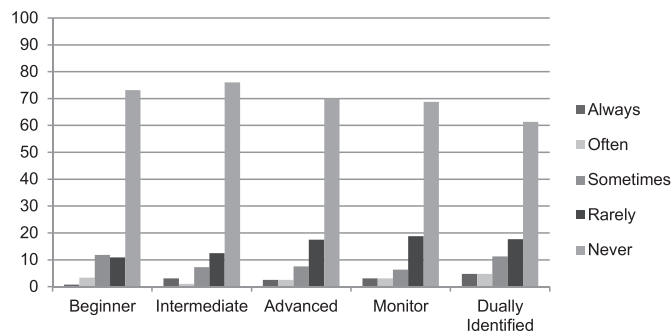


Fig. 8. How often do you provide bilingual tests for ESL students in math or science?

6. Discussion

In its focus on math and science contexts, this study has brought to light some key issues related to the nature of teacher accommodation practices for ELLs. Looking first at the identification of language proficiency levels, a compelling finding was that many surveyed teachers reported that they did not know the proficiency levels of the ELLs in their classrooms.

Though two-thirds of teachers surveyed reported uncertainty of the language proficiency levels of the ELLs in their classes; most teachers also reported that they implemented more accommodations for students with lower proficiency than higher proficiency, suggesting that many assessment practices put into place for ELLs were based upon teacher perception of students’ English ability rather than actual knowledge of proficiency levels as documented by annual WIDA administrations. Upon what these perceptions were based is unknown, and suggests a need for further research into teacher intuitions of student language proficiency.

Test exemption was also reported to take place predominantly with ELLs at the Beginner level suggesting that, despite new professional emphases in teacher education programs on inclusionary practices and differentiation for learners, some ELLs may still not be regularly included in routine classroom assessments. This may be a defensible practice, especially if the alternative is one of uniformity, which can function to be academically demotivating for ELLs in the process of acquiring English (Koretz & Hamilton, 2006). For teachers, exemption of students can also ameliorate the dilemma of allocating test grades to ELLs which may not be an accurate measure of their content knowledge; nonetheless, for students, exemption functions to exclude students from the daily rhythm of classroom life and community, and can cause students to feel further stigmatized as new students in school.

An inverse tendency was found in the survey responses between accommodations implementation and student language proficiency. This reported tendency may suggest that teachers in the sample had an understanding that systematically tapering off accommodations for ELLs is appropriate as students move toward meaningful, unaccommodated participation in classroom assessments in English. This finding may suggest that, unlike studies conducted in special education contexts (bib\_Fuchs\_et\_al\_2000b; Fuchs & Fuchs, 2001; Fuchs, Fuchs, Eaton, Hamlett, Binkley, 2000; Fuchs, Fuchs, Eaton, Hamlett, Karns, 2000; Tindal, McDonald, Tadesco, & Glasgow, 2003), there may be some level of systematicity to teachers’ decisions about accommodations implementation for ELLs. Also, this finding could suggest that teachers intuitively attend to standardization, in that as student proficiency develops, teachers tend to implement accommodations less often.

The classic, direct accommodations typically permitted on high-stakes tests were found to be variably implemented in low-stakes math and science classroom assessments. Teacher assistance and additional time were reported to be administered frequently in most classrooms, whereas translator/interpreter use, English–English dictionary use, bilingual dictionary use, and bilingual tests were not reported to be used extensively in any classroom. It is likely that practicality issues have come into play in this regard; additional time and teacher assistance can be offered during tests spontaneously and require no preparation to be implemented, whereas other accommodations, such as translations, require significant planning of resources and time which could contribute to their infrequent use (Bachman & Palmer, 1996). Further, to offer written tests in all the prospective languages of classroom test-takers is a largely impractical task; oral content translations or check-ins are much more feasible, though to do this, the teacher needs to speak the languages of the ELLs in his/her classrooms.

The reported rare implementation of native language



accommodations in this sample is likely connected to the reported monolingualism of surveyed and focal teachers. Considering that 96% of the surveyed teachers reported that they did not speak a language other than English fluently, translation of classroom test materials becomes a wholly impractical task. These teachers' linguistic abilities beyond English are not at all atypical of many small city, suburban, or rural communities in Pennsylvania, and in fact, more broadly, may be an accurate reflection of the overarching English monolingualism that is present in many educational settings in the United States (Lee, Penfield, & Buxton, 2011; Menken, 2013; Pettit, 2011).

Teachers reported rarely implementing dictionary use for ELLs on low-stakes classroom tests, and increasing dictionary use as students' English proficiency increased, suggesting that teachers believed that a certain level of English proficiency was required for use of a dictionary during assessment. Considering the meta-analysis in which Kieffer et al. (2009) found that English dictionary use was the only effective individual accommodation for ELLs on high stakes tests, the results of this study may lend support to the idea that even if ELLs have little prior experience with dictionaries from a classroom context, they may still be able to use them on high-stakes tests. More research is needed in this regard; it is a commonly accepted notion that familiarity with dictionary use is a necessary precursor for students' appropriate dictionary use on standardized measures.

The interface between the fields of special education and ESL had a consistent presence throughout this study. The finding that Dually identified students were regularly accommodated for more frequently than ELLs who did not need special education services may have two important implications: 1) IEPs continue to be powerful educational contracts between parents, students, teachers and administrators in schools, and 2) content teachers are equipped with knowledge and skills to differentiate assessment and accommodate for students' individual learning or language needs. This finding suggests that classroom accommodations implementation may be the norm for ELLs with special needs, and that these more flexible approaches toward assessment may be available for ELLs without special needs in testing situations as well.

## 7. Limitations

Several limitations to this study merit discussion here. First, as is the nature of any survey or interview process, all data reported here are self-reported and therefore, may be susceptible to a social desirability bias, or the tendency to respond in ways that participants perceive as favorable (Groves et al., 2009). In this case, teachers may have reported what they felt was appropriate practice, rather than what they actually did in assessment situations. Despite this possibility, responses were not overwhelmingly positive or negative toward accommodations implementation and the findings still yielded valuable information about teachers' reported practice during assessment.

Survey permission and distribution plans at the school district level were devised in conjunction with administrators and staff according to a strict timeline, but in the real world of schools, this timeline was not always adhered to completely, resulting in imprecision in the amount of time teachers had to respond. Distribution methods varied; in some districts, targeted techniques were used as a means of distribution directed toward math and science teachers who would participate according to provided lists. In others, the survey was sent out to a larger school listserv inviting teachers to participate more broadly, resulting in lower calculated response rates. In addition, because of the adaptive nature of the survey and the variable number of respondents to each survey question, the types of statistical analyses that could be conducted

were quite limited and therefore, the results are not widely generalizable outside of this context. Subjectivity of the Likert scale was also a limitation in that no numerical referents or examples were given to participants, resulting in participants individually interpreting and responding to the scale according to their assignments of meaning. Also, considering that such a large number of teachers reported not knowing the ELLs' WIDA levels, further report of their assessment practices for ELLs by proficiency level may lack reliability.

## 8. Implications

In an effort to progress from the “embryonic” stage of classroom accommodations research (Cizek, 2007; pg. 112), this paper begins to document what takes place in classrooms during math and science tests. It widens the scope of content assessment research for ELLs in its focus both on the classroom context and on math and science assessments, and it has multiple implications for teacher professional development and policy related to teachers of ELLs in K-12 environments.

First, the results of this study support a need for more extensive teacher professional development focusing on basic principles of assessment. Despite their high level of education across the board, participants in the study reported having taken relatively few courses and having low levels of expertise in assessment, multiculturalism, and foreign language proficiency. Critical to teachers' work with ELLs is an understanding of the importance of second language acquisition and language proficiency test scores as a valid measure of what students are capable of doing in English. Knowing and understanding students' language proficiency levels should serve as a consistent compass for teachers when assessing ELLs, and greater content teacher awareness of language proficiency levels should be fostered through teacher education and professional development.

The reported exemption of Beginner level ELLs from routine math and science tests may be an understandable and equitable practice in elementary classrooms, though implementing differentiated assessments is another way to address students' unique linguistic needs. Teachers in this study would likely benefit from more professional development focused on how to create appropriately leveled assessments for ELLs at the lowest proficiency levels in their classrooms.

More broadly, this study has implications for school and state policies related to ELLs. The finding that many accommodations had not been implemented in these classrooms to a large extent contradicts state assumptions that ELLs have previously been acquainted with the accommodations that they are offered on high-stakes tests. In fact, these results may be interpreted to mean that classroom tests provided little precedent for students' familiarity levels with accommodations in general, and native language accommodations in particular. State departments of education would be well advised to re-examine their assumptions of accommodations use when preparing test guidelines for ELLs, especially if they rely on the classroom context to facilitate high-stakes test performance.

## Acknowledgments

I gratefully acknowledge the mentorship and guidance of Dr. Lorena Llosa in the development of this study. I also thank the many school administrators and teachers who generously allowed me to conduct this research in their classrooms. This study was generously supported by a grant from The International Research Foundation for English Language Education (TIRF).

## References

- 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., Courtney, M., & Leon, S. (2003). *Research-supported accommodation for English Language Learners in NAEP*. (CSE Tech Rep. No. 586). Los Angeles: 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.
- 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.
- Acosta, B., Rivera, C., & Shafer-Willner, L. (2008). *Best practices in state assessment policies for accommodating English language learners: A Delphi study*. Retrieved January 2, 2013, from [ceee.gwu.edu/AA/BestPractices.pdf](http://ceee.gwu.edu/AA/BestPractices.pdf).
- Bachman, L., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Brindley, G. (1998). Outcomes-based assessment and reporting in language learning programmes: a review of the issues. *Language Testing*, 15(1), 45–85.
- Cho, S., & Reich, G. (2008). New immigrants, new challenges: high school social studies teachers and English Language Learner instruction. *The Social Studies Journal*, 99(6), 235–242.
- Cizek, G. (2007). Formative classroom and large-scale assessment: implications for future research and development. In J. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 99–115). New York: Teachers College Press.
- Common Core State Standards Initiative. (2013). *Implementing the common core state standards*. Retrieved January 16, 2013, from <http://www.corestandards.org>.
- Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing*, 21(3), 305–334.
- Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3), 393–415.
- Diaz-Rico, L., & Weed, K. (2002). *The cross-cultural language, and academic development handbook* (2<sup>nd</sup> ed.). Boston: Allyn and Bacon.
- Diaz-Rico, L., & Weed, K. (2009). *The cross-cultural language, and academic development handbook* (4<sup>th</sup> ed.). Boston: Allyn and Bacon.
- Ferrara, S., & DeMauro, G. (2006). Standardized assessment of individual achievement in K-12. In R. Brennan (Ed.), *Educational measurement*. Lanham: Rowman & Littlefield.
- Fuchs, L., & Fuchs, D. (2001). Helping teachers formulate sound test accommodation decisions for students with learning disabilities. *Learning Disabilities: Research and Practice*, 16(3), 174–181.
- Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C. L., Binkley, E., & Crouch, R. (2000a). Using objective data sources to enhance teacher judgments about test accommodations. *Exceptional Children*, 67, 67–81.
- Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C. L., & Karns, K. (2000b). Supplementing teacher judgments of mathematics test accommodations with objective data sources. *School Psychology Review*, 29, 65–85.
- Gottlieb, M. (2006). *Assessing English Language Learners: Bridges from language proficiency to academic achievement*. Thousand Oaks, CA: Corwin Press.
- Groves, R., Fowler, F., Couper, M., Lepkowski, J., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2<sup>nd</sup> ed.). Hoboken: Wiley & Sons.
- Harklau, L. (1994). ESL versus mainstream classes: contrasting L2 learning environments. *TESOL Quarterly*, 28(2), 241–272.
- Harper, C., & deJong, E. (2004). Misconceptions about teaching English language learners. *Journal of Adolescent and Adult Literacy*, 48(2), 152–162.
- Kieffer, M., Lesaux, N., Rivera, M., & Francis, D. (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.
- Koretz, D., & Hamilton, L. (2006). Testing for accountability in K-12. In R. Brennan (Ed.), *Educational measurement*. Lanham: Rowman & Littlefield.
- LaCelle-Peterson, M., & Rivera, C. (1994). Is it real for all kids? A framework for equitable assessment policies for English Language Learners. *Harvard Educational Review*, 64(1), 55–75.
- Lee, O. (2004). Teacher change in beliefs and practices in science and literary instruction with English Language Learners. *Journal of Research in Science Teaching*, 41(1), 65–93.
- Lee, O., Penfield, R., & Buxton, C. (2011). Relationship between "Form" and "Content" in science writing among English Language Learners. *Teachers College Record*, 113(7), 1401–1434.
- Li, H., & Suen, H. (2012). The effects of test accommodations for English Language Learners: a meta-analysis. *Applied Measurement in Education*, 25, 327–346.
- Llosa, L. (2005). Assessing English learners' language proficiency: a qualitative investigation of teachers' interpretations of the California ELD standards. *The CATesol Journal*, 17(1), 7–18.
- Llosa, L. (2007). Validating a standards-based classroom assessment of English proficiency: a multitrait-multimethod approach. *Language Testing*, 24(4), 489–515.
- Llosa, L. (2008). Building and supporting a validity argument for a standards-based classroom assessment of English proficiency based on teacher judgments. *Educational Measurement: Issues and Practice*, 47(3), 32–42.
- Llosa, L. (2012). Assessing English Learners' progress: longitudinal invariance of a standards-based classroom assessment of English proficiency. *Language Assessment Quarterly*, 9(4), 331–347.
- Maxwell, L. (2014). *Common core ratchets up language demands for English learners*. *Education Week*. Retrieved March 13, 2014 from <http://www.edweek.org/ew/articles/2013/10/30/10cc-academiclanguage.h33.html>.
- McMillan, J. (2004). *Classroom assessment: Principles and practice for effective instruction*. Boston: Allyn and Bacon.
- McNamara, T. (2001). Language assessment as social practice: challenges for research. *Language Testing*, 18(4), 333–349.
- Menken, K. (2006). Teaching to the test: how no child left behind impacts language policy, curriculum, and instruction of English Language Learners. *Bilingual Research Journal*, 30(2), 521–546.
- Menken, K. (2013). Emergent bilingual students in secondary school: along the academic language and literacy continuum. *Language Teaching*, 46(4), 438–476.
- Messick, S. (1989). Meaning and values in test validation: the science and ethics of assessment. *Educational Researcher*, 18(2), 5–11.
- Migration Policy Institute. (2015). *Number and share of English Language Learners by state*. Retrieved on September 25, 2015 from <http://www.migrationpolicy.org/programs/data-hub/charts/number-and-share-english-language-learners-state?width=1000&height=850&iframe=true>.
- National Center for Educational Statistics. (2015). *Fast facts: English language learners*. Washington DC. Retrieved September 28, 2015, from <https://nces.ed.gov/fastfacts/display.asp?id=96>.
- Oakes, J. (2005). *Keeping track: How schools structure inequality*. London: Yale University Press.
- PARCC. (2013). *Career and college readiness for the next generation*. Retrieved January 25, 2013 from <http://www.parcconline.org/sites/parcc/files/PARCC-Overview-Oct2012.pdf>.
- 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.
- Pennsylvania Department of Education. (2014). *PSSA and Keystone exams: Accommodations guidelines for English language learners*. Retrieved March 8, 2014 from <http://www.hbgisd.k12.pa.us/cms/lib/PA01000004/Centricity/Domain/38/13-14%20ESL%20Accommodation%20Manual.pdf>.
- Pettit, S. K. (2011). Teachers' beliefs about English Language Learners in the mainstream classroom: a review of the literature. *International Multilingual Research Journal*, 5(2), 143–147.
- Pitoniak, M., Young, J., Martiniello, M., King, T., Buteux, A., & Ginsburgh, M. (2009). *Guidelines for the assessment of English Language Learners*. Princeton, NJ: Educational Testing Service. Retrieved June 10, 2015 from [https://www.ets.org/s/about/pdf/ell\\_guidelines.pdf](https://www.ets.org/s/about/pdf/ell_guidelines.pdf).
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: identifying processes of classroom assessment. *Language Testing*, 18(4), 429–462.
- Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing*, 21(3), 249–258.
- Rea-Dickins, P., & Gardner, S. (2000). Snares and silver bullets: disentangling the construct of formative assessment. *Language Testing*, 17(2), 215–243.
- Reeves, J. (2004). "Like everybody else": equalizing educational opportunity for English Language Learners. *TESOL Quarterly*, 38(1), 43–66.
- Rivera, C., & Collum, E. (2006). Including and accounting for English Language Learners in state assessments. In C. Rivera, & E. Collum (Eds.), *A national review of state assessment policy and practice for English language learners* (pp. xxxiii–xviii). Mahwah, NJ: Lawrence Erlbaum Associates.
- Rivera, C., & Stansfield, C. (2004). The effect of linguistic simplification of science test items on performance of limited English proficient and monolingual English speaking students. *Educational Assessment*, 9(3–4), 79–105.
- Rivera, C., Stansfield, C. W., Scialdone, L., & Sharkey, M. (2000). *An analysis of state policies for the inclusion and accommodation of English language learners in state assessment programs during 1998–1999*. Arlington, VA: George Washington University, Center for Equity and Excellence in Education.
- Rubin, H., & Rubin, I. (2005). *Qualitative interviewing: The art of hearing data* (2<sup>nd</sup> ed.). Thousand Oaks: Sage Publications.
- Seidman, I. (2006). *Interviewing as qualitative research*. New York: Teachers College Press.
- Smarter Balanced Assessment Consortium. (2013). *School years*. Retrieved December 10, 2015 from <http://www.smarterbalanced.org/timeline>.
- Stansfield, C., & Rivera, C. (2001). Test accommodations for LEP students. *Eric Digest*, 1–6.
- Tindal, G., McDonald, M., Tadesco, M., & Glasgow, A. (2003). Alternate assessments in reading and math: development and validation for students with significant disabilities. *Exceptional Children*, 69(4), 481–494.
- US Department of Education. (2004). *No child left Behind: Testing: Frequently asked questions*. Washington, DC. Retrieved March 10, 2014, from <http://www.ed.gov/nclb/accountability/ayp/testing-faq.html#4>.
- WIDA. (2014). *WIDA access for ELLs*. Madison, WI: Center for Education Research. Retrieved June 10, 2015 from <https://www.wida.us/assessment/ACCESS/>.
- Willingham, W., Pollack, J., & Lewis, C. (2002). Grades and test scores: accounting for observed differences. *Journal for Educational Measurement*, 39(1), 1–37.
- Wolf, M., Kim, J., & Kao, J. (2012). The effects of glossary and read-aloud accommodations on English Language Learners' performance on a mathematics assessment. *Applied Measurement in Education*, 25(4), 347–374.
- Young, J., 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, 170–192.