Why Who takes What Test Matters:

The Need for Thinking Differently about English Learners and Content Tests

Tim Boals, Wisconsin Center for Education Research
Rebecca Kopriva, Wisconsin Center for Education Research
Todd Lundberg, Wisconsin Center for Education Research

Much writing about the academic success of English learners (ELs) in American schools calls attention to persistent gaps in academic achievement and opportunity: the performance of ELs on content tests continues to lag behind that of their English proficient peers, and ELs continue to have limited access to the college-bound curriculum. These gaps have come into view because growing attention on the academic achievement of ELs has been made possible by federal legislation that requires ELs to be explicitly included and accounted for in statewide assessment and accountability systems. Until the passage of the 1994 reauthorization of the Elementary and Secondary Education Act (the Improving America's Schools Act or IASA), a great percentage of English learners were routinely exempted from federal and statewide academic content tests in mathematics, language arts, and other subject areas. Before IASA, most large-scale assessment of ELs were completed at the district level and focused on literacy skills measured with English language proficiency tests (Abedi, 2007; Gottlieb, 2003). In a significant policy development, the IASA required states to develop challenging academic content standards and assessments with clearly defined performance standards and measure the growth of all students, including all ELs, against the same standards. The IASA also mandated the growth of all students be included in reports for accountability purposes, but as late as 2001 the U. S. Department of Education was still encountering inclusion problems. The next reauthorization—the 2002 No Child Left Behind Act (NCLB)—extended the accountability requirements, obligating states to report publically assessment results disaggregated by various

student groups including ELs. NCLB also required states consistently to increase the number of students scoring proficient in each of the disaggregated groups or risk withholding of funds.

The argument in the 1994 reauthorization and NCLB was that educators should be responsible for educating all of their students, not just some of them, and that they should be held accountable for demonstrating that all of their students are making progress in learning the valued content identified in the content standards. This shift in policy changed the status of ELs in American schools. Now, nearly two decades later, virtually all ELs are expected to be included in statewide mathematics and science testing in starting in their first year of attendance and in English language arts beginning in their second year. The next reauthorization will also likely require that ELs be counted prominently in accountability systems: federal pressure to teach and test ELs vis a vis challenging content is not going away.

This policy goal notwithstanding, leaders deciding how to include ELs in content assessments today still face a dilemma. The reasons for exempting ELs from full participation in statewide academic assessment systems, at least until they reached some reasonable English proficiency level, are obvious. It is nonsensical to believe that even a well-educated English speaker could meaningfully emigrate, say, to China and immediately or even in a year from arrival, show what she knows and can do on a science test written in Mandarin. The reasons in favor of including ELs in the same large-scale assessments taken by other students are as coherent, based on evidence that students left out of the accountability systems have typically been left out when educational reforms have been implemented (Abedi et al., 2005; Hakuta & Beatty, 2000; Lara and August, 1996). Inclusion legislation has greatly increased the percentage of ELs who are taking the statewide and federal achievement tests (for instance see inclusion

rates in the 2011 NAEP). For the first time policy makers are able to evaluate if progress is being made for these students as well as for other populations whose scores are disaggregated.

We want to suggest that how this dilemma is framed matters. We concede that policymakers are still hesitant about how to include ELs in academic assessment and accountability systems, and with good reason. However, we argue that new federal policy that focuses on using stopgap measures to modify how students are included in accountability systems is a step backwards. Instead, the policy should continue to insist that ELs are included *properly* in assessment systems and that their performance should be evaluated against the same unmodified achievement standards set for the general population. In particular the focus of the new federal policy should be on completing the research associated with proper inclusion and building into the policy tighter oversight mechanisms to ensure this has taken place. State and district administrators deserve to know with confidence that their information systems are providing them the best possible data in respect to how their ELs are gaining access to challenging academic curricula.

The Challenge of Fully Including ELs

The development of an appropriate framework for assessing the content knowledge, skills, and abilities of ELs is anchored in an accurate assessment of the challenge of including all students in academic assessment and accountability systems. In this short article, we focus on reframing the process by which ELs are included in the assessment of academic content connected to academic coursework in areas other than English language arts (ELA), primarily mathematics, science, and social studies. Because English language proficiency is so interwoven

with the content of ELA, especially for younger students, the challenge of assessing ELs knowledge of ELA should, we think, be taken up on it up on its own.

Appreciating the challenge of including ELs properly in assessment and accountability systems begins with attending to a set of often-repeated demographic characteristics. Children of immigrants now constitute one fifth of all U.S. school-age children. More than half of these students were born in the United States, but many come to school from U.S. households that are *linguistically isolated*, households in which no one age 14 or older speaks English exclusively or very well (Capps, Fix, Murray, Ost, Passel, & Herwantoro, 2005); 25% of these students have limited proficiency in English (Hernandez, Denton, & Macartney, 2008). National achievement data (2006-2008) show that a clear majority of states with adequate data show positive trends in the percentage of ELs scoring at or above the proficient level in reading and math (CEP, 2010). Despite these gains by ELs and significant progress on designing more accessible content assessments (see Kopriva, 2008, for a review of promising methods), EL performance on large-scale tests remains lower and varies more than non-EL student performance (CEP, 2010; Fry, 2007; Perie, Grigg, & Dion, 2005; Zhang, 2009). Underlying these statistics is a simple claim: in the 21st century, American public schools cannot be successful without providing viable schooling for linguistically diverse students (Garcia, Jensen, & Scribner, 2009).

In response to this clear need, district and school leaders have increasingly focused on ELs needs. Nonetheless, ELs continue to perform below grade level in all content areas, as reflected in significant achievement gaps reported in accountability measures (Fry, 2007; GAO, 2006; Perie, Grigg, & Dion, 2005) and are at higher risk of dropping out (Gándara & Rumberger, 2009; Rumberger, 2006; Silver, Saunders, & Zarate, 2008). While growing evidence suggests that ELs who have meaningful access to school curricula begin, after a time, to do as well as or

better than their native English-speaking peers (Flores, Painter, Harlow-Nash, & Pachon, 2009; Han & Bridglall, 2009; Kim & Herman, 2009), ELs are still often placed in classes that are remedial or do not prepare them for college, with the result that many of them fall further and further behind native English-speaking peers with the same academic capacity. These students remain trapped below academic capacity in low-level courses, or they drop out and so are not even in the schools to be counted (McNeil, Coppola, Radigan, & Heilig, 2008; Rivera & Collum, 2006). The academic assessments used to place these students and measure their progress in the content areas are typically language-intensive, leading to a false impression that the students have little knowledge or—worse yet—that they "can't learn" (Rumberger & Gándara, 2004).

Policymakers and practitioners alike now attend to the achievement gap between ELs and others and in some instances make a priority of assuring that ELs have access to core curriculum (Zehr, 2009). ELs are making progress as learners in Americans schools just as schools are making progress in serving ELs. The thorny issue that confronts standards-based educational policies is the consensus that existing assessment systems do not provide leaders and policymakers adequate information about the academic achievement of some ELs, even with test accommodations for ELs (Abedi, Hofstetter, & Lord, 2004; Emick & Kopriva, 2007; Keiffer, Lesaux, Rivera, & Francis, 2009; Solano-Flores & Li, 2006). Due largely to methodological problems that have plagued this field, the preponderance of the research on test accommodations to-date have not been able to provide adequate empirical validation for many of the accommodations. This solution can be solved in a straightforward manner by conducting experimental trials where students with identified needs are either given or not given the accommodation to address that need. While not sufficient, identifying accommodation needs by

level of English language proficiency is a good place to start. Other student variables that impact need can then be considered.

Absent adequate assessment information about these students that does not confuse content knowledge and language ability, individual schools struggle to assure students have access to appropriate academic learning opportunities and districts struggle to understand test results obtained with tests that may or may not have been accommodated properly. Under current policy, exempting ELs from tests is out of the question (see for instance Abedi et al., 2005, p. 104), and some argue *must* be out of the question (Kopriva, 2008; Kopriva & Albers, 2012). There is no clear and easy solution.

Reframing the Challenge

Leaders and policymakers alike might use this challenge as an opportunity to develop the tools they need to adequately assess their students, students who are a diverse group with widely varying language proficiency and literacy levels. This objective will require clarity about the goal of education for ELs—one group of learners in a diverse, mobile society—and the kinds of assessments that will enable ELs and those involved in their education to see progress toward that goal. First, leaders have to define the education of ELs clearly as engagement with the entire range of college- and career-preparatory, standards-aligned academic content so that ELs, like all American children, have access to opportunities to develop and build collective human capital. More than simply the opportunity to take classes, access must be to learning environments explicitly designed to elicit complex reasoning, inquiry, and metacognitive skills (Duschl, Schweingruber, & Shouse, 2007; Pellegrino, Chudowsky, & Glaser, 2001). Second, leaders have to ensure that ELs are provided with ongoing classroom assessment opportunities that reflect the

range of and depth of cognitive principles and are seamlessly integrated into rich learning activities so that teachers and students can track their true progress over time. This also means that classroom and statewide content assessment systems must include methods that enable ELs to demonstrate their ripe and ripening academic skills even though they have limited English language proficiency. Students have to be included in assessments, but the assessments have to provide students, schools, and states with information that can support valid inferences about student progress, curriculum, and resource use.

Addressing the Challenge: Two Divergent Approaches

All ELs need to be included in statewide assessment and accountability systems. There appear to be two approaches to implementing the policy. The first, by itself, grates against current policy goals. This approach focuses on a short-term solution to inadequate research for some accommodations or improper implementation of methods shown to be effective. Rather than measuring all children's academic performance by the same academic achievement standards, states might use English language proficiency to modify accountability provisions for content area achievement for ELs. The second approach has a different focus. This policy requires states to continue to measure the academic achievement with a single set of statewide achievement standards for all students except a small percentage of those with certain cognitive challenges. Both of these approaches will have significant impact on local action. We believe that the stakes are significant. Establishing separate achievement standards for any group of students, in spite of good intentions, is potentially a backdoor reversal where ELs, along with students with the most significant cognitive disabilities, would be held to lower performance. On the other hand, continuing to improve our testing systems will yield better data relative to the

same achievement standards; demonstrate our commitment to holding schools accountable for all students, not just those who it is easiest to test; and provide a model for including diverse student bodies in rich instructional opportunities.

The Same Testing and Separate Achievement Standards.

The first approach focuses on interpreting EL academic achievement against separate performance standards. It is attractive to some because this strategy obviates the pressure to complete the final research and implement testing methods that effectively minimize the barriers of language while still measuring the full range of challenging content. To be clear, we believe that those who advocate this approach see using separate achievement standards as a short-term fix until the research is completed and some standard for proper implementation has been met. They argue that, as content testing is currently conducted in most large-scale venues, this option manages to keep ELs from testing situations in which their language proficiency interferes with their opportunity to account for their content knowledge, and it protects students, families, and schools from being punished with unfair test results. Several general strategies have been suggested. An individual's language proficiency score might determine how to count her academic achievement results within an accountability system (Solorzano, 2008), or English proficiency scores might be used to set academic achievement targets for ELs that are different from those who are not ELs, or English proficiency scores might serve as a proxy for language arts achievement scores (Working Group, 2010). Cook, Linquanti, Chinen, and Jung (2012) fleshed out two of the Working Group suggestions in order to take EL's English language proficiency into account in establishing content achievement of ELs for federal accountability purposes. The first method—"progressive benchmarking"—uses time in the state school system and the ELP level of each student relative to his or her initial ELP level to adjust ELs'

achievement scores in the calculation of a school's performance. The second method—"indexed progress"— involves weighting or indexing expected English language proficiency progress with academic progress over time. Cook et al. also proposed using a growth model approach involving level of ELP and academic content growth amount that they claim do not adjust achievement scores or how they are weighted. They caution, however, that adopting a growth model methodology also presents other issues and constraints that must be attended to if used properly (Linn, 2005).

The point we wish to make is that while the general approach of adapting achievement levels or weights for accountability purposes—regardless of the method used to implement it— solves one problem, it has potentially serious unintended consequences. Schools that hold ELs to a different standard of content achievement, even if this different standard is said to be temporary until the students become more proficient in English, will likely provide them with different classroom experiences than other students. Two impacts are likely. First, when ELs' scores are less heavily weighted, there is little encouragement for schools to teach challenging content identified in the standards when this means that some type of non-language adaptations will have to be made in the classrooms for students who do not yet have the language commensurate with the complex subject matter. Truncating what ELs are taught, even for two or three years, seems to increase the probability they will fall further and further behind their English speaking peers, and increase the probability they will remain stuck and tracked into lower level classes. Second, the school experience for all learners will be diminished as the cultures and experiences of language-minority students are further marginalized.

Modifying content achievement standards or their contribution to the federal accountability system may be preferable to excluding ELs from accountability systems

altogether, but it is hard to see how this approach meets the criteria for valid and effective testing for this population. The approach clearly stops short of assuring that each state or consortium develops an assessment system that produces accurate measurements of all students' academic achievement. More, this approach potentially distracts policymakers from the real goal: ongoing, adequate teaching of challenging grade level content for all students in American schools.

21st-Century Testing and Common Achievement Standards.

Rather than waiting to assess fully ELs academic achievement until they have learned enough academic language to be assessed on the full range of challenging content, this approach argues that the policy goal remain consistent: states and consortia are obligated to measure all students except those with significant cognitive deficits against the same, unmodified achievement standards that are weighed the same in the accountability systems; policymakers remain committed to completing necessary research and establishing implementation and oversight standards to ensure that ELs get what they need during testing. This approach, we suggest, leads ultimately to making valid inferences about the content knowledge, skills and abilities of ELs with confidence. The focus of this approach is on the continued improvement of content testing so as to provide all learners with effective and valid opportunities to reveal their content area knowledge relative to the same content unmodified achievement targets to which their native English speaking peers are held accountable. The improvement of content will include better accommodations research and implementation and wider range of testing strategies.

Better accommodations research and implementation. In the past decade, the nature of the question concerning effective assessment of ELs has begun to change (Kopriva, 2008). Pennock-Roman and Rivera (2011) are the latest in a series of educators and researchers to call

for rigorous experimental research designs, and designs that investigate accommodations effectiveness by English language proficiency levels. Prior studies have tended to focus on post hoc analyses of performance data where no randomizing of student conditions took place. Further, almost all of the accommodations research classified English learners as a single group even though it is well known that EL student needs systematically vary by a number of characteristics, most centrally how well they can communicate in academic English. Evaluating the usefulness of accommodations by splitting ELs into groups by English proficiency level and possibly other characteristics such as literacy in students' first academic language, would greatly improve the interpretability of the experimental designs.

Fidelity of proper accommodation assignments for English learners taking content tests and consistent implementation of those assignments has been a huge problem. For instance, ongoing efforts at systematizing NAEP inclusion criteria for both ELs and students with disabilities across states have been dogged by different interpretations of the criteria. Douglas (2004) reported that teachers of ELs had little understanding of how to assign large-scale test accommodations even though general state guidelines were available, and in their review of literature, Koran, Kopriva, Emick, Monroe, and Garavaglia (2006) found that teacher or committee assignments are notoriously inaccurate in providing the same accommodations for EL students with similar profiles. Additionally, Koran et al. (2006) found that teacher accommodation choices proved to be no better than randomly generated choices, even when they were told exactly what to focus upon to make their decisions. These problems led to the development of a computerized instrument called STELLA that assigns test accommodations for English learners based on their individual needs. A randomized study found that this approach led to significantly improved test scores for students receiving proper accommodations, while

students receiving improper accommodations performed no differently than those not receiving any accommodations at all, even though they needed them (Kopriva, Emick, Jipolito-Delgado, & Cameron, 2007). Approaches such as STELLA—approaches that are able to take the needs of language learners into account in a consistent manner across students with similar profiles — need to be implemented with oversight to ensure the students are assigned and are receiving what is at least minimally acceptable.

   Better assessment forms. For a number of students with language and literacy challenges, including lower English proficient ELs, using accommodations on top of standard test forms is not enough to provide consistent access on tests especially when the tests measure complex content. The kinds of language that serve as the primary vehicle for mainstream academic instruction (Bruna & Gomez, 2008; Schleppegrell, 2004) create a hurdle for students lacking facility in comprehending and producing the distinctive academic language registers associated with challenging content. A substantial body of literature has demonstrated that such students benefit from instructional approaches using multiple modes of representation (see for example Echevarria, Vogt, & Short, 2007; Schleppegrell, 2004; Wright, 2008). Such approaches have also been found beneficial with native English-speaking students lacking the types of literacy skills needed to access the language-intensive text of textbooks, classroom activities, and assessment tasks of all types (Bruna & Gomez, 2008), and recent work in assessment research is calling attention to types of student content achievement that can be represented in non-linguistic or multi-semiotic modes (see for example Bateman, 2008; Kress & van Leeuwan, 2006). Underlying the use of alternative representations—such as performance, relational techniques, and interactive manipulation of stimuli—is the assumption that they can be directly analogized to

the linguistic means traditionally used to communicate knowledge and skills and that they relate similarly to other cognitive machinery (Gee, 1999; Schleppegrell, 2004; Talmy, 2003).

One promising method is a technology-interactive, dynamic approach called ONPAR. This methodology measures content, including complex content typically associated with more sophisticated language and literacy structures, utilizing a range of multi-semiotic techniques to primarily convey meaning while significantly reducing the language load. To-date a series of ONPAR projects have developed prototypes in elementary and middle school science and mathematics and in high school end-of-course biology and chemistry. Randomized trials have compared how focal groups and control students have performed on ONPAR tasks compared to traditional tasks measuring the same content at the same level of cognitive complexity.

The elementary and middle school science study focused on low English proficient students and found that these students performed significantly better on ONPAR than traditional forms, while non-ELs scored similarly on the two forms. Further, controlling for science ability, the low English proficient ELs performed as well as their non-EL peers. This erasure of the achievement gap is an important milestone. Results also suggest that test makers might consider that this approach be used for select students in such a way that they can be considered comparable to results taken by other students on tests using traditional formats. A recently completed mathematics study focused on how students with learning disabilities, other students with disabilities teachers thought might benefit, and struggling native English speaking readers performed on ONPAR forms compared to traditional forms and compared to non-ELs with no IEPs. Taking into consideration the mathematics ability of the students, findings were similar for students with learning disabilities as what was found in the science study. Students with speech and language, attention deficit, emotional and autistic disabilities showed similar promise

although the n's were too low to conduct inferential analyses, and analysis of how struggling readers fared is beginning. The data collection of the high school prototypes is currently underway.

Conclusion

Meeting the challenge of accounting for learning in diverse schools requires school leaders to enlarge their thinking about what it means to account for content knowledge. Mislevy's (1994) Evidence Centered Design work has already laid the theoretical groundwork for testing under varying conditions. As assessment comes to be understood as methods of using analogs between language and other representations, we argue that properly designed and implemented flexibility in assessment conditions might also be seen as the transformation of language-intensive forms and tasks into occasions that are more accessible to those lacking the language, literacy, or related challenges without jeopardizing the targeted construct at its intended level of cognitive complexity. We encourage policy makers to resist depending on measures of academic language proficiency to recalculate results for accountability purposes. Better policy can create incentives for a next generation of better accommodations research and implementation strategies, as well as test forms and tasks that enable test makers adequately to decouple language from content that, executed appropriately, can provide us a truer window into students' content knowledge, skills, and abilities. The policy goal needs to be allowing test takers to demonstrate adequately their knowledge, conceptual, reasoning, and inquiry skills using thoroughly vetted and standardized options, all while retaining the fidelity of the results relative to the same set of achievement standards. Test developers now have recourse to approaches to collecting achievement information that are sensitive to both linguistic and nonlinguistic

representations of content and the relations that link these representations to cognitive associations. Moreover, completing the accommodations research and pursuing consistent and viable assignment and implementation strategies are within our grasp.

We want to be very clear that we do not mean to suggest that students do not need to learn academic language. High-level knowledge building in any discipline happens through the discourse of the discipline; all students need an opportunity to learn academic languages and research into the relationship between academic language proficiency, student background, opportunities to learn, and content assessments will remain critical. As ELs become proficient in academic registers, the nature of how they take content tests should change. Ultimately, the students' content knowledge should be tested through the discourse—language structures as well as other forms of representation—typically used to make meaning within the relevant academic discipline.

In an increasingly diverse society, however, we argue that students who cannot yet adequately show what they know and can do via traditional text-based methods need to be included in the full range of content schooling in ways that are accessible in the classroom, and they need to be in fully included in assessments that are able to reflect their knowledge, skills, and abilities at complex as well as basic levels. To achieve the latter, we need to keep improving our testing and holding test makers, educators, and researchers responsible for their roles in bringing this to fruition. The leaders of a diverse democracy must, we argue, continue to press for and insist upon holding ELs to the same content and achievement standards to which all other students (except a small number of students with severely cognitive disabilities) are held. They must press for and insist upon defensible, accessible assessment, improving upon a base of literature that, to-date, has at least defined the issues and begun to address them. The way

forward is through rigorous empirical study, improved testing decisions that are carried into practice, and not becoming distracted from this goal, just as directions for next steps are becoming clear.

References

Abedi, J. (Ed.). (2007). *English language proficiency assessment in the nation: Current status and future practice*. Davis: University of California.

Abedi, J., Bailey, A., Butler, F., Castellon-Wellington, M., Leon, S., & Mirocha, J. (2005). *The validity of administering large-scale content assessments to English language learners: An investigation from three perspectives*. (CSE Technical Report 663). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

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.

Bateman, J. (2008). *Multimodality and genre: A foundation for the systematic analysis of multimodal documents*. Basingstoke, England: Palgrave Macmillan.

Bruna, K., & Gomez, K. (Eds.) (2008). *Talking science, writing science: The work of language in multicultural classrooms*. Mahwah, NJ: Erlbaum.

Capps, R., Fix, M., Murray, J., Ost, J., Passel, J., & Hermwantoro, S. (2005). *The new demography of America's schools: Immigration and the no child left behind act*. Washington, DC: Urban Institute.

Center on Education Policy. (2010). *State test score trends through 2008-09* (Parts 1-5). Retrieved from the Center On Education Policy website: http://www.cep-dc.org/

Cook, H. G., Linquanti, R., Chinen, M., & Jung, H. (2012). National evaluation of Title III implementation supplemental report: Exploring approaches to setting English language proficiency (ELP) performance criteria and monitoring English learner progress. . Retrieved from http://www2.ed.gov/rschstat/eval/title-iii/implementation-supplemental-report.pdf

Douglas, K. (2004) *Teacher ideas on teaching and testing English language learners: Summary of focus group discussions (C-SAVE Rep. No. 223)*. College Park, MD: University of Maryland, Center for the Study of Assessment Validity and Evaluation (C-SAVE).

Duschl, R., Schweingruber, H., & Shouse, A. (Eds.). (2007). *Taking science to school: Learning and teaching science in grades K–8*. Washington, DC: National Academies Press.

Echevarria, J., Vogt, M. E., & Short, D. (2007). *Making content comprehensible for English learners: The SIOP model* (3rd ed.). Boston, MA: Pearson Allyn & Bacon.

Emick, J. E., & Kopriva, R. J. (2007). *The validity of large-scale assessment scores for ells under optimal testing conditions: Does validity vary by language proficiency?* Paper presented at the American Educational Research Association annual meeting.

Fry, R. (2007). *How far behind in math and reading are English language learners?* Washington, DC: Pew Hispanic Center. Retrieved from http://pewhispanic.org/files/reports/76.pdf

Gándara, P., & Rumberger, R. W. (2009). Immigration, language, and education: How does language policy structure opportunity? *Teachers College Record, 111*(3), 750–782.

Garcia, E. E., Jensen, B. T., & Scribner, K. P. (2009). The demographic imperative. *Educational Leadership, 66*(7), 8-13.

Gee, J. (1999). Reading and the new literacy studies: Reframing the National Academy of Sciences Report on Reading. *Journal of Literacy Research*, *31*(3), 355–74.

Gottlieb, M. (2003). *Large-scale assessment of English language learners: Addressing educational accountability in K-12 settings.* Alexandria, VA: Teachers of English to

Speakers of Other Languages.

Government Accountability Office. (2006). *No Child Left Behind Act: Assistance from education could help states better measure progress of students with limited English proficiency* (Report GAO-06-815). Washington, DC: Author.

Flores, E., Painter, G., Harlow-Nash, Z., & Pachon, H. (2009). *Que pasa?: Are English language learning students remaining in English learning classes too long?* Los Angeles, CA: Tomas Rivera Policy Institute.

Hakuta, K. & Beatty, A. (2000). *Testing English language learners in U.S. schools: Report and workshop summary*. Washington, DC: National Academy Press.

Han, W. J., & Bridglall, B. L. (2009). Assessing school supports for ELL students using the ECLS-K. *Early Childhood Research Quarterly, 24*(4), 445–462.

Hernandez, D. J., Denton, N. A., & Macartney, S. E. (2008). Children in immigrant families: Looking to America's future. *Social Policy Report, 22*(3), 3-22.

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.

Kim, J., & Herman, J. L. (2009). A three-state study of English learner progress. *Educational Assessment, 14*, 212–231.

Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). London, NY: Routledge.

Kopriva, R.J. (2008). *Improving testing for English language learners*. New York: Routledge.

Kopriva, R.J. & Albers, C. (in press). Considerations for testing special needs students. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology.* Washington, DC: American Psychological Association.

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., Emick, J., Monroe, J. R., & Garavaglia, D. (2006, April). Teacher and multisource computerized approaches for making individualized test accommodation decisions for English language learners. Paper presented at the annual meeting of the National Council of Measurement in Education, San Francisco, CA.

Lara, J. & August, D. (1996) *Systemic reform and limited English proficient students*. Washington, DC: Council of Chief State School Officers.

Linn, R.L. (2005, December). Adjusting for differences in tests. Paper for a symposium on the use of school-level data for evaluating federal education programs, Washington, DC: The Board on Testing and Assessment, The National Academies.

McNeil, L. M., Coppola, E., Radigan, J., & Heilig, J. V. (2008). Avoidable losses: High-stakes accountability and the dropout crisis. *Education Policy Analysis Archives, 16*(3), 1–48.

Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika, 58*, 79–85.

Pellegrino, J., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

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 & Practice, 30*(3), 10-28.

Perie, M., Grigg, W., & Dion, G. (2005). *The nation's report card: Mathematics 2005* (NCES 2006-453). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2005/2006453.pdf

Rivera, C., & Collum, E. (Eds.). (2006). *A national review of state assessment policy and practice for English language learners*. Hillsdale, NJ: Erlbaum.

Rumberger, R. (2006, October). Tenth grade dropout rates by native language, race/ethnicity, and socioeconomic status. *ELL Facts, 7*. Retrieved from Linguistic Minorities Research Institute website: http://www.lmri.ucsb.edu/publications/elfacts-7.pdf

Rumberger, R. W., & Gándara, P. (2004). Seeking equity in the education of California's English learners. *Teachers College Record, 106*(10), 2,032–2,056.

Schleppegrell, M. (2004). *The language of schooling: A functional linguistics perspective.* Mahwah, NJ: Lawrence Erlbaum Associates.

Silver, D., Saunders, M., & Zarate, E. (2008). *What factors predict high school graduation in the Los Angeles Unified School District?* Retrieved from the California Dropout Research Project website: http://cdrp.ucsb.edu//pubs_reports.htm

Solano-Flores, G., & Li, M. (2006). The use of generalizability (g) theory in the testing of linguistic minorities. *Educational Measurement: Issues and Practice, 25*(1), 13-22.

Solorzano, R. W. (2008). High stakes testing: Issues, implications, and remedies for English language learners. *Review of Educational Research, 78*(2), 260-329.

Talmy, L. (2003). *Toward a cognitive semantics*. Boston, MA: MIT Press.

Working Group on ELL Policy. (2010). Improving educational outcomes for English language learners: Recommendations for the reauthorization of the Elementary and Secondary Education Act. Retrieved from http://ellpolicy.org/wp-content/uploads/ESEAFinal.pdf

Wright, L. J. (2008). Learning by doing: The objectification of knowledge across semiotic modalities. *Linguistics and Education, 19*, 225–243.

Zehr, M. (2009). Adpoted tongue: English-learners pose policy puzzle. *Education Week, 28*(17), 8-17.

Zhang, Y. (2009). *Conflicts between state policy and school practice: Learning from Arizona's experience with high school exam policies*. Retrieved from the Center On Education Policy website: http://www.cep-dc.org/