



The Gordon Commission  
on the Future of Assessment in Education

**Assessment of Content and Language in  
Light of the New Standards:  
Challenges and Opportunities for English  
Language Learners**

---

*Kenji Hakuta*  
*Stanford University*

---

Paper prepared for the Gordon Commission on the Future of Assessment. I would like to acknowledge the contributions of Patricia Gandara and Virginia Gonzalez who provided reviews of an earlier draft of the paper. This paper also draws from the work of colleagues from Understanding Language: <http://ell.stanford.edu>

This paper plays out an imagined scenario in 2017 (5 years hence) for the assessment of English language learners, based on assumptions about what I know of the Common Core State Standards and how this most recent wave of reform will impact state and local systems in the assessment of content and English language proficiency.

### **Current State**

The current state is as follows. The Common Core State Standards in English Language Arts and Mathematics has been adopted by 48 states. Two consortia of states (PARCC and SBAC) are developing assessments aligned to these standards for implementation by 2014. The National Research Council published a framework for K-12 science (National Research Council, 2012), which is being developed into the next generation science standards by Achieve, and many states have committed to joining in its development, and will very possibly adopt these new standards.

What is already clearly evident in the substance of the Common Core as well as the next generation science standards is the prominence of language -- not just language as in parts of speech, grammar and vocabulary, but also high levels of language embedded in the instructional actions expected in the new standards. It is the aspects of language that linguists would call pragmatics and systemic functional linguistics. They expect language to support the following sorts of academic expectations:

“Students can, without significant scaffolding, comprehend and evaluate complex texts across a range of types and disciplines and they can construct effective arguments and convey intricate and multifaceted information” (ELA student portraits, p. 7)

Mathematically proficient students understand and use stated assumptions, definitions, and previously established results in constructing arguments. They make conjectures, and build a logical progression of statements to explore the truth of their conjectures” (Math practices, pp. 6-7)

This new angle on the language demands of the new standards has begun to be documented by the Understanding Language Initiative (ULI – available at [ell.stanford.edu](http://ell.stanford.edu)) and the papers are being widely disseminated, with the primary audience being educators working on the content areas who are coming to the realization that these content-embedded aspects of language are operational units of what has been traditionally and broadly termed “academic language”.

The new standards, while clearly driving assessment through the federally-funded consortia of states (PARCC and SBAC), have also begun to shape other parts of the system attempting to align to the standards – state and district policies regarding curriculum, professional supports, textbooks and materials, and teacher preparation. My own observation in having engaged in various conversations with players and providers in these efforts to align the new standards systemically is that they vary all over the map. The variation is best characterized as piecemeal and unsystematic on one end, to strategic and thoughtful on the other. Piecemeal and unsystematic efforts will result in very little reform, whereas strategic and thoughtful gives it a chance. In turn, this variation is primarily attributable to system capacity to take on such a profound fundamental instructional shift that is called for by the new standards.

In addition to the content assessments based on the Common Core and the new science standards, states are currently required to develop a separate assessment of English Language Proficiency (ELP) that “corresponds” to the state content standards. English Language Learners are required by federal law to be identified, classified, and provided with additional appropriate services until their educational needs arising from their language minority status have been addressed, and they are reclassified as “English proficient”. In addition, the ELP assessments are used for Title III accountability purposes under current NCLB. There is a strong enough constituency base for Title III assessments so that these provisions will remain in any reauthorization of ESEA. I do not expect any changes in the federal mandate for identification and for special services, required under *Lau v. Nichols* (1974) to change, and thus a need for a “screener” assessment would remain to identify this class of students.

The current state of formative assessment is also worthy of note. If you ask most school district officials, the response to the question, “What is formative assessment?” would basically describe some small version of the annual summative assessment, rather than as something that is part of instructional practice. But there is increasing traction of the latter view. Much credit for this emerging shift goes to Margaret Heritage (2010) who has been a powerful advocate for

effective formative assessment practice, as having the following three components: (1) teacher adjustment of instruction in response to assessment evidence; (2) students receiving feedback about their learning; and (3) student participation in the process through self-assessment (Black & William, 1998). What is relevant for language is the fact that this interactive, instructionally-based perspective of formative assessment can only be successful with effective uses of language by both the student and the teacher. And if this practice is supported effectively, it can be a powerful agent of change in instruction by providing ELL students with a rich and powerful source of language input for their English language development in the context of content instruction.

The view of language is also undergoing considerable transformation. The traditional view of language as primarily forms made up of structures, rules, and vocabulary has given way to the addition of functions that serve general cognitive and pragmatic goals. More recent views characterize language as embedded in a set of domain-specific actions within larger systems of meaningful discourse within the classroom (van Lier and Walqui, 2012).

The present state of education for ELLs has a number of important issues worth noting. One is the problem of addressing the needs of long-term English language learners. A second is the status of the native language in instruction. A third issue is recognition of the advantages of bilingualism.

First, the requirements of NCLB to report disaggregated data on ELLs has called direct attention to the problem of long-term English learners, often defined as those students who began schooling in the U.S. but remain in ELL status after more than 5 years (Olsen, 2010). The problem of long-term ELLs is seen as a failure of moving the students past oral English proficiency and reading fluency into academic literacy, a term that has come to be called “academic language”. In many school districts, almost half of the ELL students who start in elementary school are not reclassified as English-proficient within the 5-6 year time frame. The problem of moving ELLs whose academic progress has stalled has become a major preoccupation of school districts, in large part due to NCLB accountability requirements. The state of California passed legislation in 2012 requiring school districts to identify, report, and address the needs of long-term ELLs (AB 2193).

Second, the question of bilingual education is an age-old one, but one that continues to be debated. The evidence to its relative effectiveness compared to English-only programs is well-

established (Goldenberg, 2008), yet it continues to stir deep emotions about linguistic nationalism and even anti-immigration sentiments. I have argued elsewhere that this debate is a distraction from paying attention to other important instructional issues (Hakuta, 2011). The debate unfortunately demoralizes bilingual teachers and certainly discourages innovation, but it has been difficult to move beyond political positioning. However, the anti-bilingual forces may have ebbed from its peak of the post-Proposition 227 mania that started in California and travelled across the country. But the symbolic politics of language, immigration, and easily-stirred sediments of xenophobia continue to lurk in the background, all this in spite of all the rhetoric about globalization.

The third issue, about the value of bilingualism, can be witnessed in the popularity of the dual-language programs that aim for bilingualism for both the language-minority and the native English participants in the program. These are widely popular in large part because of the participation of middle-class Anglophones in the program who see the advantages of bilingualism (these are the people who accept and even embrace the inevitability of globalization). The benefits of these programs for ELLs are not well documented but widely presumed. Better-documented are the positive benefits of bilingualism for various aspects of executive function and for long-term life benefits such as the delay in onset of dementia (Bialystok 2010, 2011; Bialystok, Craik & Freedman, 2007), consistent with much of the advances in our understanding of human learning and cognition (National Research Council, 2000). The popularity of this idea among the middle-class can be seen by the fact that two New York Times articles that appeared in the past two years on this research attained the “most e-mailed” status each for several consecutive days [Dreifus, 2011; Bhattacharjee, 2012]. Thus, the idea of bilingualism as a benefit has legs, though perhaps still only for the middle-class.

### **Predictions for 2017**

For expository purposes, I will write this section in the present tense for 2017, in order to avoid as much as possible uses of the future perfective (“the new tests will have been in place for three years”). So, it is now 2017, and we can now look back over the past 5 years. The new assessments aligned to the Common Core in English Language Arts/Literacy and Mathematics, after a glitch due to an uneven technology infrastructure in many schools, have now entered their third year of administration. As of this year, these tests have been augmented with an assessment

of the next generation Science Standards. Furthermore, after some controversy, standards in history/social studies have been developed, and in several years, they will be added to the list of subjects covered.

A very positive change occurred as the assessment system transitioned into the new standards by recognizing the fact that the specification of the content area constructs necessitates explicit recognition of their language demands. The Understanding Language Initiative<sup>1</sup> was explicit in identifying the key shifts that began occurring as early as 2012, when districts and states started to align their work with the Common Core. In the summary paper from their conference, they reported the following key shifts, and the language challenges and opportunities contained therein:

**Shifts in ELA:**

1. Reading: reading and comprehending literature and informational texts of increasing complexity to build knowledge
2. Writing: using evidence to inform, argue and analyze for varied audience/purposes and present knowledge gained through research
3. Speaking and Listening: Working collaboratively, understanding multiple perspectives, and presenting ideas
4. Language: choosing language and conventions to achieve particular functions, purpose, and rhetorical effects

**Opportunities in ELA:**

1. Reading: Leverage background knowledge, build strategic competence, and provide supports to allow access to texts rather than simplifying or “pre-empting” the text.
2. Writing: Draw upon background strengths to develop content for writing and scaffold writing itself; Provide ELs with meaningful engagement with mentor texts, including opportunities to focus on language and text structure; Ensure that writing is meaningful communication.

---

<sup>1</sup> See papers available through the Understanding Language website at <http://ell.stanford.edu>

3. Speaking and Listening: Provide opportunities for extended discourse & engagement with academic texts; Develop meaningful collaborative tasks that allow students to use their full linguistic/cultural resources; Teach ELs strategies to engage with text in multiple ways.

**Shifts in Mathematics:**

1. mathematical practices provide opportunities for students to engage in posing and solving problems, explaining concepts and making connections, understanding multiple representations of mathematical concepts and models, communicating their thought processes through procedures, justifying reasoning, and making arguments.
2. Instruction should support mathematical discussions and use variety of participation structures (teacher-led, small group, pairs, student presentations, etc.) that allow students to use multiple representations (diagrams, charts, symbols, models, etc.) in communicating math content and practices.

**Opportunities for math:**

1. Focus on students' mathematical reasoning, not language proficiency
2. Shift to a focus on mathematical discourse practices, move away from simplified views of language
3. Recognize and support students to engage with the complexity of language in math classrooms
4. Treat everyday language and experiences as resources, not as obstacles

**Shifts in Science:**

1. Inquiry redefined as a set of eight practices including four sense-making practices that are particularly language intensive and parallel to similar demands in math and ELA.
2. Focus on a limited set of core concepts, and on building understanding coherently over multiple years of school.

### Opportunities in Science:

1. Immersion in science content through observation, investigation and discourse provides a language learning opportunity
2. Models and visual representations of information as used in science are a resource and bridge for language learners to grasp content as they develop language.
3. Science text and discourse are new to all students, so attention to the language challenges inherent in them supports science learning and language development for all students.

Realization of the ubiquitous nature of language in the content areas have led to several important contributions to the field:

1. *An operational definition of “academic language”*. Prior to 2012, attempts to define academic language consisted of efforts removed from the academic practices of the disciplines. For example, academic language was contrasted against “conversational language” or “contextualized language” (Cummins, 1981; Snow, 1987); defined as a set of words that appear across academic content texts (Coxhead); or defined as sentence frames and language routines that could be used in content language classrooms (Dutro & Kinsella, 2009). With a focus on the language demands of the shifts in the content area standards, academic language was operationalized as those language demands embedded in the instructional actions necessary to meet the content shifts. By bringing to the foreground the nature of academic language embedded in the classroom practices of the content areas, the assessments are now able to report separately on the performance of students on an “academic language” dimension within the content areas.
2. *Language proficiency for all students*. The ability to measure academic language within the content tests also built a bridge between the language needs of ELLs and native speakers of English, because all students must develop proficiency in the use of language in the content areas. This eliminated the old problem of setting an arbitrary “cutoff score” for ELLs to be reclassified as proficient in English, because their performance can be directly placed in the performance range of native English speakers. Furthermore, measurement of academic language for all students also meant that language proficiency could be monitored for English language learners throughout their academic career, while under the old system, former ELLs reclassified as proficient were no longer monitored for their English language proficiency.

3. *Sharper focus on English language proficiency needs of ELLs.* This content-based operational definition of language proficiency also helped to define what was meant by “correspondence” between English language proficiency standards and the new content standards. This was a concept that came about through regulatory actions by the Obama administration in providing waivers to states from compliance with the AYP (Adequate Yearly Progress) requirements under NCLB. In that memo, one requirement for the waiver was for the state to have in place state ELP standards aligned to the state content standards. Most states applied for the waiver, and then developed (through the WIDA consortium and through other state collaborative endeavors) ELP standards utilizing a framework for ELP standards created through the leadership of the Council of Chief State Schools Officers. This framework essentially divided the ELP standards into a two-stage model, in which the first stage paid attention to students in their beginning stages of English language development, and the second stage elaborated on performance expectations on explicit language demands within anchor standards in the content areas. Adoption of this new framework for ELP standards helped sharpen the focus of the ELP assessments, focusing their efforts on basic levels of English performance and general academic language performances, while the content-embedded language assessments were moved to the content assessments.

4. *Increased appreciation of good formative assessment practice as language support.* During the past 5 years, the field has come to a realization that formative assessment practice is good instructional practice rather than a good measurement tool, as advocated by Heritage (2011). Given the realization that language shapes how content knowledge is shaped, taught, and used for demonstrating understanding, formative assessment practices have keyed in to the role of helping the teacher understand the relationship between the student’s readiness to engage with the language of the content being taught. Savvy teachers have appropriated the many technology tools made available by advances in the field of natural language processing to support these practices.

Another major development that was only fantasized in 2012 is the availability of massive amounts of text in digital form, including all textbooks available on the new computing platforms through publishers, Apple’s iBooks initiative, and open source materials through efforts by the Creative Commons. Furthermore, integration of multimedia resources and the revolution in voice recognition technology has blurred the traditional distinctions between

written and spoken language. Searching through texts and tags on a virtually infinite world of resources in multiple languages is now a reality, and part of the “Big Data” problem of computer science<sup>2</sup>. We can now envision student products of learning being continuous multimedia portfolios of cognitive and interactive representations (interactions with peers, teacher, texts and outside expertise) that can be linguistically measured, and used for purposes of formative assessment as well as depicting individual student as well as group-level longitudinal learning trajectories.

Now in 2017, just a few years into the Common Core and related new assessments, we are in the midst of improving the first-generation system put into place by the assessment consortia. Because of the rush to implementation, those assessments only achieved part of the original vision. They fell particularly short on fully realizing the content-embedded language dimensions as a new construct, and paid more attention to the more traditional aspects of language (grammar, vocabulary, and generic academic functions) than to the uses of language reflecting content-specific acts. They also fell short on using the formative assessment opportunities in a creative manner to encourage teacher-student and student-student engagement. And they also fell short on measuring oral language skills due to the inability to leverage voice recognition and NLP technology adequately.

The assessments, however, have gained broad traction in focusing on core academic skills and driving home to all educators the realization that language is an issue for all students, with an additional portion of English language proficiency necessary to address the development of newcomer second language learners.

### **Opportunities in 2016**

As we look to the second-generation assessments around the common core and related subject areas, the following opportunities present themselves:

1. A more sophisticated content-embedded language construct based on the experiences of the first implementation of CCSS to enable better reporting of a separate subscore on language capabilities for content domains or subdomains, reported for *all* students. The addition of language scores for all students would recognize the common challenges of the language of

---

<sup>2</sup> New York Times, April 7, 2012, Berkeley Group Digs In to Challenge of Making Sense of All That Data.

academic content for all students and bring this understanding to the attention of all educators, to break out of the categorical mentality that has helped to bring special attention to ELLs, but also served as an excuse to release the “mainstream” educator from responsibility for ELLs.

Additionally, this will better serve the accountability purposes of ESEA in keeping track of ELLs for whom continued growth in English language proficiency remains a concern, because in the present system, those who attain English proficiency and therefore are reclassified as fluent are no longer tracked and drop out of the accountability system<sup>3</sup>.

2. A targeted assessment system for ELLs that ramps up to the content-embedded language construct, but that defines various instructional and learning functions of language looking at the needs of beginning ELLs; this assessment would yield an annual score, but should be part of an on-going assessment that looks at multimedia portfolios created by students, which can be evaluated using NLP methodologies. The present system for assessing English language proficiency creates a low-reliability / high stakes situation, especially in the early elementary grades when most ELLs are initially identified and enter the system (i.e., excluding those students who are newcomers to the country in the later grades). That is to say, the ELP assessments are often used to identify and classify students in kindergarten when students are least familiar and comfortable with formal testing, and once they are classified, there are programmatic consequences so long as they remain an ELL (until they go through a formal reclassification process that become more rigorous with each successive grade). Once classified, students receive targeted services which are for the benefit of the students, but given the unreliability of the initial assessment, students may or may not actually need these services, and as students get older, many resent the academic stigma of being in non-mainstream programs. An assessment that is continuous throughout the year would avoid putting all the eggs in the annual assessment for classification and reclassification purposes. An assessment that gives a picture of growth during the year is also well-suited to the particular nature of the ELL category which, unlike categories such as race or disability, is fundamentally developmental in nature, and rapid changes are expected.

3. Dynamic formative assessment practices that are supported by NLP technologies that actively absorb linguistic materials around the content, enable multimedia and multi-party

---

<sup>3</sup> See recommendations of the Working Group on ELL Policy at <http://ellpolicy.org>

engagement with the content, and offer feedback to teachers and students on the next steps of learning; importantly, these assessment practices allow students and teachers to engage with language and texts in the world outside of the immediate classroom. This component might be imagined as the sampling of the language in the multiple facets of learning and relating it, through computational NLP resources, to similar replications that can be found in the world outside the classroom – imagine a multidimensional Google search that includes texts, images, music, videos, widgets etc. that come organized as a microcosm of knowledge and learning and adapted to the classroom environment. Students are asked to engage with these representations, and their interpretations would give teachers clues to adapt instruction, and also give students an opportunity to assess their own understanding as well as collaborate with their peers (these three components correspond to the principles of effective formative assessment practice as formulated by Heritage, 2010).

4. Bilingual assessments through authentic translation tasks -- many Americans have now recognized the broadly positive values of student ability in multiple languages, as evidenced by the growing numbers of two-way bilingual programs, immersion programs in non-English languages, and biliteracy programs that build on the native language of language minority students. Translation and interpretations abilities will become a premium task that exemplify the advantages of bilingualism. For students in the growing number of programs that value bilingualism, translating knowledge representations from one language to the other (and critiquing the quality of translations as a “comprehension” version of the same ability) becomes a highly authentic assessment. Much of the evaluation of the quality of translations can be done by machine, as can item generation following algorithms that mimic good and faulty translations, and therefore these assessments can be tightly targeted with respect to relevance to the particular curriculum.

### **Needed: Epistemological Linguistics**

These new developments will require good theory and strong engineering solutions. In order to support the development of assessments as envisioned here, a new field of applied scholarship that pays explicit attention to the relationship between language and content will need to be cultivated, in order to attract the best human talent available to maximize the potential. Traditionally, this is represented in the interdisciplinary field broadly known as

language and cognition, encompassing the fields of linguistics, cognitive psychology, sociocultural theory, philosophy, computer science, and the neurosciences. I am proposing basically an applied version of that field, call it *epistemological linguistics*, that is unabashedly interdisciplinary and applied to understand how language operates in the context of knowledge domains (hence epistemology) in different contexts and systems of learning.

The goal of epistemological linguistics are:

- to *describe* language in the context of epistemological domains.
- to *understand* the role of language in learning, individual knowledge representation, and the social life of the knowledge in classrooms and beyond.
- to *explain* the causal connections that may exist between enriched language representations and enhanced cognitive, sociocognitive and sociocultural exchanges of knowledge domains.

Examples of language behaviors by students might include:

- Describing different forms of knowledge representation (e.g., narrative, perspective, visual, historical, causal...).
- Constructing a domain-specific argument and participate in content-building discourse (e.g., mathematical, scientific, historical argument).
- Using formulaic phrases and making stylistic / sociolinguistic moves that are traditional to content-specific genres.
- Making lexical choices and connecting these choices to knowledge networks.
- Actively using NLP technologies in engaging with epistemologies and texts.

Examples of questions that studies in such a field might address include:

- How does turntaking, discourse participation, engagement in virtual communities, etc. serve as a foundation for knowledge development?
- What key commonalities and differences emerge from structural, functional, and systemic analyses of canonical texts used in different knowledge domains?
- What level of lexical analysis is necessary in order to adequately account for vocabulary effects in content area learning?

- What are the best ways to characterize teacher language behaviors in order to account for variations in student learning of different knowledge areas?
- What are the aspects of language that are critical to knowledge representation and what are peripheral, so that tests can contain more of the critical and less of the peripheral aspects of language?
- How can we understand cross-language transfer of knowledge to support students who are in dual language programs?

Appropriate assessment of content-embedded language given the sophisticated tools made available through technology poses a fascinating and important challenge for the future of assessment, and it is important to get our best minds to work on all angles of this complex problem.

## References

- Bhattacharjee, Y. (2012); Why bilinguals are smarter. New York Times, March 17, 2012.
- Bialystok, E. (2010). Bilingualism. *Wiley interdisciplinary reviews: Cognitive science*, 1, 559-572.
- Bialystok, E. (2011). Coordination of executive functions in monolingual and bilingual children. *Journal of Experimental Child Psychology* 110, 461-468.
- Bialystok, E., Craik, F. & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia*, 45: 459-464.
- Black, PJ & Wiliam, D. (1998), Assessment and Classroom Learning. *Assessment in Education: Principles, Policy and Practice*, 5, 7-73.
- Council of Chief State Schools Officers (2012). Framework for English Language Proficiency Development Standards corresponding to the Common Core State Standards and the Next Generation Science Standards. Washington, DC: CCSSO.
- Coxhead, A. (2000). An Academic Word List ELI Occasional Publications #18. University of Wellington, English Language Institute.
- Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In California State Department of Education (Ed.). *Schooling and language minority students: A theoretical framework*. (pp. 3-49). Los Angeles: National Dissemination and Assessment Center.
- Dreifus, C.(2011). The bilingual advantage. New York Times, May 30, 2011
- Dutro, S. & Kinsella, K. (2009). English Language Development: Issues and Implementation at Grades 6-12, Chapter 3 of Improving Education for English Learners: Research-Based Approaches. Sacramento, CA: California Department of Education.
- Goldenberg, C. (2008). Teaching English Language Learners – What the research does – and does not say. *American Educator*, Summer (2008) 8-44.
- Hakuta, K. (2011). Educating language minority students and affirming their equal rights: Research and practical perspectives. Seventh Annual Brown Lecture in Education Research. *Educational Researcher*, 20: 1-12.
- Heritage, M. (2010) Formative Assessment and Next-Generation Assessment Systems: Are We Losing an Opportunity? Washington, DC: Council of Chief State School Officers.

National Research Council (2000). *How people learn: brain, mind, experience, and school*. Committee on Developments in the Science of Learning and Committee on Learning Research and Educational Practice, Commission on Behavioral and Social Sciences and Education, National Research Council. Washington, DC: National Academies Press.

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Olsen, L. (2010). *Reparable Harm: Fulfilling the Unkept Promise of Educational Opportunity for California's Long Term English Learners*. Long Beach: Californians Together.

Snow, C.E. (1987). Beyond conversation: Second language learners' acquisition of description and explanation. In J. Lantolf & A. Labarca (Eds.), *Research in second language learning: Focus on the classroom* (pp. 3-16). Norwood, NJ: Ablex.

Van Lier, L. & Walqui, A. (2012) *Language and the Common Core State Standards*. Commissioned paper for Understanding Language. Downloaded at <http://ell.stanford.edu/publication/4-language-and-common-core-state-standards>