



KELLIE ROLSTAD¹ - *rolstad@umd.edu*

JEFF MACSWAN² - *macswan@umd.edu*

NATALIA GUZMÁN³ - *nlguzman@umd.edu*

University of Maryland

Artículo recibido: 06/07/2014 - aceptado: 28/10/2014

BILINGUAL LEARNERS AND THE PURPOSES OF LANGUAGE ASSESSMENT

ABSTRACT:

US educational policy, first under No Child Left Behind (NCLB, 2001) and then under the Common Core State Standards Initiative (CCSSI, 2010), introduced a narrow definition of language proficiency. This definition validates the language variety of social elites and excludes the varieties of other socioeconomic groups. Although the official definition of language proficiency in current federal and state policies recognizes traditional conceptions of bilingualism, we argue that it fails to recognize *sociolinguistic status* (Rolstad, 2005) or socioeconomically-marked language differences. As a result, currently

¹ Served as associate professor of Language and Literacy and of Early Childhood Education at Arizona State University before joining the faculty of the University of Maryland. Her research interests include language education, language diversity, and multilingualism, in addition to recent work on democratic and alternative education. She has published numerous articles and book chapters; examples of her work appear in *Educational Policy*, *Teachers College Record*, *Bilingual Research Journal*, *Bilingual Review*, and *Hispanic Journal of Behavioral Sciences*, among other venues.

² Professor of Applied Linguistics/TESOL in the Department of Teaching and Learning, Policy and Leadership, and affiliate professor of Linguistics at the University of Maryland. MacSwan's applied research program is focused on the role of language in theories of school achievement and on education policy related to bilingual learners in US schools; his basic scientific research program is concerned with the linguistic study of bilingual codeswitching. He is the author of over fifty publications, an editor of the *International Multilingual Research Journal*, and a fellow of the National Education Policy Center.

³ Doctoral student in Second Language Education and Culture in the College of Education at the University of Maryland. She completed a bachelor's degree at the University of Buenos Aires. In addition, she holds a master's degree in Hispanic Linguistics as well as in Teaching English as a Second Language from the Pennsylvania State University. Her academic interests include foreign and second language education, classroom-based research, and language assessment. She is also interested in second language acquisition and applied linguistics with a focus on heritage language learners. Natalia has taught Spanish for several years in Argentina and the U.S.

avored language assessments may tend to misidentify bilingual or English-proficient children as limited English language learners. US schools use language assessment for the dual purpose of initial identification and reclassification; we argue that initial identification tests should assess language across sociolinguistic strata, while reclassification tests should be more tied to school-based language (Mahoney & MacSwan, 2005a; MacSwan & Rolstad, 2006). Implications for language assessment in multilingual communities are drawn.

KEYWORDS: Language assessment, Common Core State Standards Initiative, Bilingualism, English language learners.

RESUMEN:

La política educativa de los Estados Unidos, primero bajo la ley Que Ningún Niño Se Quede Atrás (NCLB, 2001) y luego bajo la Iniciativa de los Estándares Comunes Estatales (CCSSI, 2010), introduce una definición limitada de competencia lingüística. Esta definición valida la lengua de élites sociales y excluye las variedades lingüísticas de otros grupos socioeconómicos. Si bien la definición oficial de competencia lingüística en leyes federales y estatales actuales reconoce concepciones de bilingüismo tradicionales, sostenemos que fracasa en reconocer el *estatus sociolingüístico* (Rolstad, 2005) o las diferencias lingüísticas marcadas socioeconómicamente. Como resultado, las evaluaciones de lengua favorecidas actualmente tienden a identificar erróneamente a niños bilingües o competentes en inglés como estudiantes con habilidad limitada en la lengua inglesa. Las escuelas estadounidenses usan las evaluaciones de lengua con el doble propósito de identificación inicial y reclasificación; sostenemos que las evaluaciones de identificación inicial deben evaluar el lenguaje a través de estratos sociolingüísticos, mientras que las evaluaciones de reclasificación pueden estar más conectadas al lenguaje de la escuela (Mahoney & MacSwan, 2005a; MacSwan & Rolstad, 2006). Nuestras conclusiones tienen consecuencias para la evaluación lingüística en comunidades multilingües.

PALABRAS CLAVE: Evaluación lingüística, Estándares Comunes Estatales, bilingüismo, estudiantes aprendices del idioma inglés.

1. BILINGUAL LEARNERS AND THE PURPOSES OF LANGUAGE ASSESSMENT

US educational policy, first under No Child Left Behind (NCLB, 2001) and subsequently under the Common Core State Standards Initiative (CCSSI, 2010), introduced a narrow definition of language proficiency which validates the language variety of social elites to the exclusion of the language varieties of other socioeconomic groups. As such, while the official definition of language proficiency implicit in current federal and state policies may recognize traditional conceptions of bilingualism, it fails to recognize socioeconomically-marked language differences, or what Rolstad (2005) called *sociolinguistic status*. As a result, currently favo-

red language assessments are likely to misidentify bilingual or English-proficient children in school as having limited English language proficiency. Recognizing that language assessments for bilingual learners⁴ in US schools are used for the distinct purposes of initial identification and reclassification, we argue that tests used for initial identification be designed to assess language across sociolinguistic strata, while tests used for reclassification might be more tightly tied to school-based language (Mahoney & MacSwan, 2005; MacSwan & Rolstad, 2006). The US is not alone in managing the pressures of national education policies affecting minority students (Hornberger, 2009), and there is mounting interest in promoting assessment-driven instruction around the world (LMTE, 2013). The issues related to the Common Core State Standards (CCSS) addressed here pertain now to the US context, but may be part of policy debates internationally in the near future.

2. LANGUAGE PROFICIENCY IN NCLB AND THE COMMON CORE STATE STANDARDS

The creation of NCLB in 2001 brought important changes to US educational policy, in particular in relation to bilingual learners and language in school. As Menken (2009) indicates, NCLB «offers a striking instance of high-stakes educational testing functioning as de facto language policy» (p. 103) that has emphasized a shift towards English-only instruction since it requires that all students, including bilingual learners, take the same state academic content assessments in Reading/Language Arts, Math, and Science. Menken (2009) notes that «NCLB terminated Title VII, the Bilingual Education Act, which had been part of the Elementary and Secondary Act (ESEA) since 1968, and that the word *bilingual* has now been entirely expunged from the legislation» (Menken, 2009, p. 105; also see Crawford, 2004b, and Evans & Hornberger, 2005). Title III of NCLB, «English Language Acquisition, Language Enhancement, and Academic Achievement Act», focuses on bilingual learners and their language development. The first purpose of instruction listed by Title III of NCLB is to «ensure that children who are limited English proficient, including immigrant children and youth, attain English proficiency, develop high levels of academic attainment in English ...» (Title III, sec. 3102). Bilingual learners are characterized as «children who are limited English proficient», which not only positions students as members of a remedial group but also reinforces a deficit orientation towards bilingual learners (Wiley, 2013). In

⁴ We use the term bilingual learners to refer to school-age bilinguals for whom both languages may be fully proficient or one language may still be developing (English, for sequential bilinguals in the US context). We will use this term rather than *English Language Learners (ELLs)* or Limited English Proficient students (LEPs) except where context requires otherwise.

addition, as Crawford (2004b) highlights, the «English Language Learner» group itself is a problematic construct in NCLB's accountability system since it constitutes a highly diverse population in terms of «socioeconomic status, linguistic and cultural background, level of English proficiency, amount of prior education, and instructional program experience» (p. 3). These characteristics of bilingual learners are not specifically taken into account when NCLB holds all students in all schools accountable to the same standard.

Concerning the role of the student's first language, Title III of NCLB does not make any reference to other languages besides English. Even the term «limited English proficient» underscores the importance of the target language while concealing the student's first language background. There are no statements regarding the value of multilingualism or a role for the student's first language in the community or the school. Evans and Hornberger (2004) note that in NCLB the student's native language is not even recognized for its facilitative role in promoting English language development: «Indeed, it may be viewed as a crutch in subject area study that prevents children from making adequate progress toward English language proficiency» (p. 89). Worse, English language is viewed as the indispensable condition for students' academic achievement and Title III emphasizes the development of children's English language proficiency in conjunction with academic achievement. Not surprisingly, Evans and Hornberger (2004) conclude that NCLB reflects not only a language-as-problem orientation (Ruiz, 1984) but also a monolingual view of bilingual learners regarding the development of language and literacy skills.

The CCSS follow in the spirit of NCLB to create a national common set of academic standards. The National Governors Association Center for Best Practices (NGA Center) and the Council of Chief State School Officers (CCSSO) coordinated the CCSS to develop K-12 standards for English Language Arts and Math that were released in 2010. The CCSS does not include separate English language development/proficiency standards for bilingual learners, although it expects bilingual learners to be held to the same levels of English Language Arts standards, with some acknowledgement of the different learning experiences of ELL students, like in the case of NCLB. (See TESOL International Association, 2013, for further discussion.)

It is important to note that the CCSS developers later recognized the importance of bilingual learners and their language needs in a brief addendum («Application of Common Core State Standards for English Language Learners») to the general guidelines. In contrast to Title III of NCLB, this document recognizes English Language Learners (ELLs) as a «heterogeneous group with differences in ethnic background, first language, socioeconomic status, quality of prior schooling, and

levels of English language proficiency» (CCSSI, 2010, p. 1). It also acknowledges that «ELLs bring with them many resources that enhance their education and can serve as resources for schools and society» (p. 1), including their first language, literacy knowledge and skills. Although the document notes the importance of students' first language in instruction, the emphasis on the acquisition of English remains, as in NCLB.

These policy developments led to increased interest in English language proficiency standards and corresponding assessments of language proficiency in the states. In addition, the CCSS developers released an *English language proficiency development* (ELPD) framework to assist states in revising their standards to bring them into conformity with the ELPD. Bailey and Butler (2003) assert that English language assessments need to «capture the necessary prerequisite language proficiency for mainstream classroom participation and for taking content-area assessments in English» (p. 1), which Bailey and Wolf (2012) note has become «a general consensus» in current policy contexts (p. 1).

The «general consensus» emerged at least in part because NCLB required each state adopt an English language proficiency test statewide, and that the test be aligned to the English language proficiency standards, which in turn must be aligned to the content areas (Kuriakose, 2011). In actual practice, then, federal law required that bilingual learners' language proficiency be assessed through the lens of school language, blurring content and language, whereas language assessments had previously been developed in connection with a theory of language proficiency (albeit poorly; see MacSwan, Rolstad & Glass, 2002; MacSwan & Rolstad, 2006; MacSwan & Mahoney, 2008).

3. LANGUAGE, CONTEXT, AND SCHOOLING

In the context of the CCSS and related ELPD framework, the construct of Academic Language Proficiency requires a careful and meaningful definition, especially if it is to be properly assessed in schools (Wiley & Rolstad, 2014). However, despite its significance and centrality in the current policy context, the construct itself remains extraordinarily elusive. As Wong Fillmore and Snow (2002) pointed out,

Although there is a lot of discussion about the need for all children to develop the English language skills required for academic learning and development, few people can identify exactly what those skills consist of or distinguish them from general Standard

English skills. To the extent that this matter is examined at all, observers have usually pointed to differences between written and spoken language (p. 27).

This observation is echoed in Snow and Uccelli (2009) and Uccelli and Meneses (this issue).

In order to evaluate or measure language in any group, we must first clearly understand what language is as a construct (Gee, 2014; MacSwan & Rolstad, 2010). Research on language acquisition has found cross-culturally that all typically developing children acquire the language of their respective speech communities, and do so effortlessly and without instruction (Pinker, 1994; Slobin & Bowerman, 1985). During the most active acquisition period in the preschool years, children learn approximately 10-12 new words per day, often on one exposure and in highly ambiguous circumstances (Gleitman & Landau, 1994), and acquire knowledge of elementary aspects of sentence structure for which they have no evidence at all (Lightfoot, 1982). Moreover, as Tager-Flusberg (1997) has pointed out in a review of the literature, «by the time children begin school, they have acquired most of the morphological and syntactic rules of their language» (p. 188) and possess a grammar essentially indistinguishable from adults. These facts and others have led researchers to believe that language acquisition is inwardly directed by innate principles of Universal Grammar (Chomsky, 1981; Berwick, Pietroski, Yankama, & Chomsky, 2011), or an internal «bioprogram», as Bickerton (1981) termed it.

Universal Grammar (UG) is presumed to be a biological endowment common to the human species, and unique in essential respects. It defines a narrowly delineated hypothesis space for the language learner, who uses primary linguistic data from the speech community to set options permitted by UG. Thus, UG begins in an initial state, S_0 , which successively approximates the language of the speech community through a series of intermediate states, $S_1 \dots S_n$, until it reaches a steady state, S_S , after which it appears to undergo only peripheral changes.

Of course, languages differ across communities of speakers, and across individuals as well (Fillmore, Kempler & Wang, 1979). Thus, when we identify a speech community as «speakers of English» or «speakers of Tyrolian German» we engage in an idealization, assuming homogeneity for the purpose of discussion, much in the way that natural scientists assume homogeneity of body organs and other objects of study. We might usefully think of «speakers of a language L» as those speakers whose languages are each sufficiently alike as to permit intelligible intercommunication in L. In doing so, however, we recognize that in actual fact speech communities have considerable internal variation, even to the level of individual speakers (*idiolects*), and that speakers may be members of multiple speech communities.

As Chomsky (1995) noted, a language is «a state of the language faculty [which] is some accidental product of varied experience ...» (p. 7). Formally, a particular language - such as English, German, or Swahili - is a set of expressions defined by a grammar, a psychological mechanism which maps sound to meaning and which is represented in the mind/brain of a speaker-hearer, and a vocabulary. The grammar of a particular language is a set of values over the range of variation permitted by UG once it has entered the steady (or «mature») state. In the context of first language (L1) acquisition, then, we take «language proficiency» to be a state of linguistic maturity in which a learner has acquired a grammar which is compatible with the language of the community of origin (MacSwan & Rolstad, 2010).

Literacy and other school subjects will no doubt make use of a child's language ability, but these seem substantially different in character. Humans acquire their first languages by instinct, upon exposure, the way birds acquire birdsong; but the learning of school subjects, such as literacy, physics and mathematics, does not follow a biologically endowed program (Chomsky, 1986; Gee, 2001; MacSwan & Rolstad, 2010). Academic achievement denotes a domain of knowledge that is specific to a particular human context - namely, the world of schooling. While all (typical-developing) children develop a vocabulary and a grammatical system, not all children will come to know specific facts about geography, history or physics.

In addition, a child's tacit understanding of the rules which govern language use is also sensitive to social and situational contexts, and the interpretation of particular linguistic expressions is tied to a language user's appreciation of relevance, coherence, and context (Sperber & Wilson, 1986). Language use has the effect of establishing a *who* and a *what*, a socially - situated person engaged in a particular kind of craft or activity, in Gee's (1999) terms - a teacher, doctor, member of a club or street gang, a regular at a local bar, or a student at school. Gee (1999) uses the term «social language» to denote the role of language in Discourse, the set of conventions that result in an expression of personal and social identity, and of relationships among interlocutors and participants. Thus, as we each make meaning out of language, we do far more than compute an interpretation deriving from the interaction of syntax and word meaning. We make use of context, a wealth of knowledge and theories about the world, and of a particular set of cultural models, practices, and beliefs.

Thus, *language* reflects a grammatical system which consists of the rules and principles which govern syntax (word order), morphology (word formation), and phonology (pronunciation), and which makes use of context to interface with principles of discourse, pragmatics, and semantic interpretation. Speakers and communities differ with regard to the particular form these rules and principles might take,

resulting in the formation of distinctive varieties and conventions on language use; but each community nonetheless has a language every bit as rich and complex as the next (Crystal, 1986; Newmeyer, 1986; Milroy & Milroy, 1999). With regard to second language acquisition among school-age children, we naturally expect to see a maturational process which proceeds on an independent timetable, with ongoing evidence of development in core linguistic systems. (See MacSwan & Pray, 2005, for further discussion.)

Central to these points about the nature of linguistic knowledge is the observation that languages vary within and across communities, and are likely to differ according to the context in which they were acquired. Some varieties have higher social prestige than others, but the prestige associated with a linguistic variety results from social and political forces which are altogether independent of the linguistic system itself. Linguistic prescriptivists made the error of assuming that the language variety of the formally educated presented certain cognitive advantages underlying the relative social and political success of the groups represented, an assumption which made its way into educational theories about the academic achievement of minorities in US schools (MacSwan & Rolstad, 2010; Wiley & Rolstad, 2014).

Gee (1986), Wiley (2005), and Wiley and Rolstad (2014) criticized the widespread view that a cognitive Great Divide separates literate and nonliterate people, and that the acquisition of literacy has powerful, positive effects on cognitive development. The notion of a Great Divide is further evident in early tensions surrounding efforts by structuralist linguists to document the «linguistic equality» of language varieties, which had serious consequences for sacred distinctions that kept privilege in the hands of the educated classes (Newmeyer, 1986). Clearly, any language can be a suitable vehicle for thought and learning, and any typical child who speaks a language already has all that is needed to achieve full cognitive development. Indeed, this basic idea explains the finding that children do better academically when taught bilingually (MacSwan & Rolstad, 2005, 2010; Rolstad, Mahoney & Glass, 2005; Krashen, Rolstad & MacSwan, 2012).

A widely recognized early and influential perspective on Academic English derives from Cummins' (1979) distinction between *Basic Interpersonal Communication Skills* (BICS) and *Cognitive Academic Language Proficiency* (CALP). In important respects, the BICS/CALP distinction serves as the foundation of the CCSS focus on Academic English - not only for bilingual learners, but also for English-speaking children (TESOL International Association, 2013; Anstrom, DiCerbo, Butler, Katz, Millet, & Rivera, 2010). Cummins distinguished the two in developmental terms:

In monolingual contexts, the [BICS/CALP] distinction reflects the difference between the language proficiency acquired through interpersonal interaction by virtually all 6-year-old children and the proficiency developed through schooling and literacy which continues to expand throughout our lifetimes. For most children, the basic structure of their native language is in place by the age of 6 or so but their language continues to expand with respect to the range of vocabulary and grammatical constructions they can understand and use and the linguistic contexts within which they can function successfully (Cummins, 2000a, p. 63).

Similarly:

In short, the essential aspect of academic language proficiency is the ability to make complex meanings explicit in either oral or written modalities by means of language itself rather than by means of contextual or paralinguistic cues such as gestures and intonations (Cummins, 2000a, p. 59).

One sees a clear attribution of special linguistic status to the language of the educated classes. CALP is said to involve the ability to make meanings explicit by means of language itself rather than by means of gestures and intonations. Of course, there is no reason to believe, and no evidence to support, the presumption that academics are better at explaining their craft than the less-schooled are at explaining theirs, or that accompanying gestures are less useful to academics than to others. Imagine a typical professor of English, for instance, trying to talk in detail about farming, boatbuilding, or auto repair. Academics would typically lack knowledge of relevant vocabulary in these contexts - words which would be «low frequency» for them, but not for many others.

Cummins (2000b) also viewed BICS and CALP as different with regard to how much «knowledge of language» is involved in each:

Considerably less knowledge of language itself is usually required to function appropriately in interpersonal communicative situations than is required in academic situations. ... In comparison to *interpersonal conversation*, the *language of text* usually involves much more low frequency vocabulary, complex grammatical structures, and greater demands on memory, analysis, and other cognitive processes (p. 35-36, emphasis added).

One sees an important category error in Cummins' discussion of the differential attributes of BICS and CALP. CALP is defined as «the ability to make complex meanings explicit in either *oral* or *written* modalities», in contrast to BICS, in which meaning making is substantially aided «by means of contextual or paralinguistic cues such as gestures and intonations» (Cummins, 2000a, p. 59, emphasis added). Gestures are typically associated with oral communication rather than written language. In Cummins (2000b), written language seems to be identified with CALP and oral language with BICS, even though oral language can also have the features

of CALP. There are, of course, many non-academic varieties of writing which one could readily compare with academic abilities to determine whether they might differ with regard to «the ability to make meanings explicit», and one could similarly compare oral language in academic and professional contexts with oral language in other contexts to determine whether one variety is more dependent on gestures and paralinguistic cues than the other. Regrettably, no such empirical analysis is offered by proponents of the distinction.

Writing and speaking have well known distinctive characteristics in their own right; owing to the absence of phonological representation, writers rely on punctuation and word choice, while speakers have the advantage of using intonation, pitch, stress, and pauses to add texture to their message. Perhaps most importantly, speech is typically spontaneous and subject to non-linguistic interference (distraction, fatigue, forgetting), while writing is usually planned and edited, possibly over the course of a considerable timespan. Comparing the language of text in school contexts with oral language in out-of-school contexts clearly introduces considerable confusion into any effort to clarify the nature of Academic Language.

Others have sought to illuminate the characteristics of Academic English in more recent work, attempting to extend and differentiate their approach from Cummins' (1979) original BICS/CALP distinction. For instance, Chamot and O'Malley (1994) defined Academic English as «the language that is used by teachers and students for the purpose of acquiring new knowledge and skills ... imparting new information, describing abstract ideas, and developing students' conceptual understanding» (p. 40). Bailey and Huang (2011) characterize Chamot and O'Malley's (1994) definition as imprecise, noting that «it still does not provide sufficient specificity of exactly 'what' language is being used in school» (p. 350). Schleppegrell (2001), in contrast, defines Academic English as a special register that entails a specific «constellation of lexical and grammatical features» used in school contexts (p. 431). More concretely, Schleppegrell asserts that school-based language contains a high degree of structure, «realized through elaboration of noun phrases, sentence rather than prosodic segmentation, and clause-structuring strategies of nominalization and embedding» (p. 431). Schleppegrell is a systemic functionalist whose approach emphasizes the contextual attributes of language use, following Halliday (1961). Schleppegrell's discussion is therefore focused on academic language as a locus of language use. Even so, a limitation of Schleppegrell's definition is that it appears to be mainly based on written texts that circulate in school, such as pedagogical materials and student writing samples. The implication is that school-based genres are mostly written, while spoken genres characterize the language outside school, since «children's out-of-school experiences provide them with many more opportunities for interactional language use» (pp. 437-438). This assumption

overlooks the use of spoken language within school, and reduces language outside of school to oral interaction alone. The BICS/CALP distinction also seems to permeate Schleppegrell's definition; school-based texts are characterized as «specific», «technical», and «dense», while spoken interaction is defined as «generic» and «sparse» (p. 438).

A common and uncontroversial theme in the academic English literature is the realization that academic English is associated with a particular context, namely, school, whereas out-of-school language is associated with other contexts. This is a simple observation, true by definition. However, a very controversial claim, either stated overtly or implied by omission, is that the language of school has special status and capacity, not shared by language used in most other contexts, such that it enables its speakers to thrive intellectually in ways which the language of other contexts could not (see MacSwan & Rolstad, 2005, 2010; Wiley, 2005; Gee, 2014; Rolstad, 2014; Wiley & Rolstad, 2014). We have discussed the latter supposition, and argued that it is unfounded, following from traditional language ideologies and confusion introduced by inappropriate comparisons of written and spoken language.

4. THE PURPOSES OF LANGUAGE ASSESSMENT

Since the passage of the Bilingual Education Act (1968) and *Lau v Nichols* (1974), schools have been required to determine whether newly enrolling children might need language-related assistance at school (Crawford, 2004a). In a survey of state practices, Mahoney and MacSwan (2005b) found that the majority of states (79 %) used a home language survey for this purpose, while others used teacher observations, parent referral, administrative recommendation, or referral from previous school. If the home language survey or other method indicates that a language other than English is used at home, then the newcomer's language is assessed for the purpose of *initial identification* as an English Language Learner. Later in a child's program of study, language is assessed as part of the *reclassification* process.

It is important to underscore these two distinct assessment purposes. Initial identification is intended to determine whether the child may have a *language barrier*, which might make participation in the classroom very difficult or impossible. There may be other difficulties the child faces, such as the effects of limited prior schooling or poverty; these are important factors, but they are distinct from language proficiency. With regard to initial identification, we are interested in knowing

whether a child has acquired English before coming to school; since the context of acquisition might not be school, we reasonably expect that the child's English may be colored by non-school contexts. Thus, assessment for initial identification should be guided by a theory of language structure and acquisition; it should look to assess language proficiency regardless of the context in which it was acquired.

Assessment for *reclassification* assumes very difficult language learning circumstances; in this case, the context of language acquisition is school. After entering a program of study specifically designed to teach children English along with school content, children are assessed to determine whether their English has developed to a level of proficiency permitting them to meaningfully participate in schooling - that is, whether they can use English alone to further develop their knowledge of school subject matter. Thus, assessment for reclassification should be guided by a theory of language structure and acquisition, and should target language as it is specifically used in school contexts; it should also include an assessment of English literacy.

Differentiation of assessment according to its specific purposes is consistent with Messick's (1994) influential work focused on the validity of the use of test scores, now codified in the AERA/APA/NCME Standards (2014). Messick (1989) articulated a unified concept of validity with multiple facets: «Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment» (Messick, 1989, p. 13). Hence, validity encompasses «... the interpretability, relevance, and utility of scores, the import or value implications of scores as a basis for action, and the functional worth of scores in terms of social consequences of their use» (Messick, 1989, p. 13).

Hence, language assessments which do not properly differentiate language assessment of bilingual learners for the distinct purposes of *initial identification* and *reclassification* are likely to result in erroneous interpretations of test scores and negative consequences for children. For instance, a «language test» constructed according to a set of school standards, intended as outcome measures of instruction, will likely wildly misidentify children as English Language Learners if used for the purpose of initial identification.

Correct identification and reclassification are important to achieve, and using a portfolio of multiple indicators is likely to assist in achieving greater accuracy (AERA/APA/NCME, 2014; Mahoney & MacSwan, 2005a). Incorrect initial identification means that a child who may need language assistance services will not receive them, or that one who does not need such services, but may have other needs, will be incorrectly identified. If we err in reclassification, it may mean that children

will stop receiving services prematurely; if we exit children too late, they may spend more time in ESL programs than is likely to be beneficial, as Callahan, Wilkinson and Muller (2010) warn.

5. SUMMARY

Language assessments for bilingual learners in US schools are used for the different purposes of initial identification and reclassification; however, under the influence of NCLB, CCSS, and the ELPD framework, language assessments have come to rely less on traditional language theory and much more on conformity to performance standards, which in turn were developed to promote what proponents call Academic English (that is, English used in schools). We have argued that tests used for the purpose of initial identification should be designed to assess language across sociolinguistic strata, as pre-NCLB language proficiency tests attempted to do, and should detect proficiency regardless of the context in which the language was acquired, while tests used for reclassification might be more tightly tied to school-based language.

REFERENCES CITED

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education [AERA/APA/NCME]. (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Anstrom, K., DiCerbo, P., Butler, F., Katz, A., Millet, J., & Rivera, C. (2010). *A Review of the literature on Academic English: Implications for K-12 English Language Learners*. Arlington, VA: The George Washington University Center for Equity and Excellence in Education.
- Bailey, A. & Wolf, M. (2012). The challenge of assessing language proficiency aligned to the Common Core State Standards and some possible solutions. Paper presented at the Understanding Language Conference at Stanford University, Stanford, California.
- Bailey, A., & Butler, F. (2003). An evidentiary framework for operationalizing academic language for broad application to K-12 education: A design document. Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California, Los Angeles.
- Bailey, A., & Huang, B. (2011). Do current English language development/proficiency standards reflect the English needed for success in school? *Language Testing*, 28, 3, 343-365.
- Berwick, R., Pietroski, P., Yankama, B., & Chomsky, N. (2011). Poverty of the Stimulus Revisited. *Cognitive Science*, 1-36.
- Bickerton, D. (1981). *The roots of language*. Ann Arbor, MI: Karoma Press.
- Bilingual Education Act (1968). Pub. L. No. (90-247), 81 Stat. 816 (1968).
- Callahan, R. & Wilkinson, L. & Muller, C. (2010). Academic achievement and course taking among language minority youth in US schools: Effects of ESL placement. *Educational Evaluation and Policy Analysis*, 20(10), 215-229.
- Chamot, A., & O'Malley, J. (1994). *The CALLA handbook: Implementing the cognitive academic language learning approach*. Reading, MA: Addison-Wesley.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris Publications.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.
- Chomsky, N. (1995). *The minimalist program*. Cambridge: MIT Press.
- Common Core State Standards Initiative [CCSSI]. (2010). *K-12 Common Core State Standards*. Retrieved from CCSSI Website: www.corestandards.org/the-standards.
- Crawford, J. (2004a). *Educating English Learners: Language Diversity in the Classroom*. 5th edition. Los Angeles: Bilingual Education Services.
- Crawford, J. (2004b). No Child Left Behind: Misguided approach to school accountability for English language learners. Paper presented at the Forum on Ideas to Improve the NCLB Accountability Provisions for Students with Disabilities and English Language

- Learners. Washington, DC: Center on Education Policy & National Association for Bilingual Education.
- Crystal, D. (1986). The prescriptive tradition. In D. Crystal (Ed.), *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.
- Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, 49, 222-251.
- Cummins, J. (2000a). Putting language proficiency in its place: Responding to critiques of conversational/academic language distinction. In J. Cenoz & U. Jessner (Eds.), *English in Europe: The acquisition of a third language*. Clevedon, UK: Multilingual Matters.
- Cummins, J. (2000b). *Language, Power and Pedagogy: Bilingual Children in the Crossfire*. Clevedon, UK: Multilingual Matters.
- Evans, B., & Hornberger, N. (2005). «No Child Left Behind: Repealing and unpeeling federal language education policy in the United States». *Language Policy*, 4, 87-106.
- Fillmore, C., Kempler, & Wang, S. (1979). *Individual differences in language ability and language behavior*. New York: Academic Press.
- Gee, J. (1986). Toward a realistic theory of language acquisition. *Harvard Educational Review*, 56 (52-68), 1986 [Reprinted in Masahiko Minami & Bruce P. Kennedy, Eds., *Language Issues in Literacy and Bilingual/Multicultural Education*, Reprint Series No. 22, Cambridge, Mass: Harvard Educational Review, 1991, pp. 98-116]
- Gee, J. (1999). *An introduction to discourse analysis: Theory and method*. New York: Routledge.
- Gee, J. (2001). Progressivism, critique, and socially situated minds. In C. Dudley-Marling & C. Edelsky (Eds.), *The fate of progressive language policies and practices*. Urbana: National Council of Teachers of English.
- Gee, J. (2014). Decontextualized language: A problem, not a solution. *International Multilingual Research Journal*, 8(1), 9-23.
- Gleitman, L., & Landau, B. (1994). *The acquisition of the lexicon*. Cambridge: MIT Press.
- Halliday, M. A. K. (1961). Categories of the theory of grammar. *Word*, 17(3) 241-92.
- Hornberger, N. (2009). Multilingual education policy and practice: Ten certainties (grounded in Indigenous experience). *Working Papers in Educational Linguistics*, 24(2), 1-18.
- Krashen, S., Rolstad, K. & MacSwan, J. (2012). Review of «Research summary and bibliography for Structured English Immersion programs» of the Arizona English Language Learners Task Force, pp. 107-119. In C. Faltis & B. Arias (eds.) *English Learners in Arizona*. Bristol, UK: Multilingual Matters.
- Kuriakose, A. (2011). *The factor structure of the English Language Development Assessment: A Confirmatory Factor Analysis*. Doctoral dissertation, Arizona State University.
- Lau v. Nichols (1974). 414 U.S. 563, 94 S.Ct. 786 (1974).
- Lightfoot, D. (1982). *The language lottery: Toward a biology of grammars*. Cambridge: MIT Press.

- Learning Metrics Task Force [LMTF]. (2013). *Toward universal learning: Recommendations from the Learning Metrics Task Force*. Montreal and Washington, D C: UNESCO Institute for Statistics and Center for Universal Education at the Brookings Institution.
- MacSwan, J. & Mahoney, K. (2008). Academic bias in language testing: A construct validity critique of the IPT I Oral Grades K-6 Spanish Second Edition. *Journal of Educational Research and Policy Studies*, 8(2), 85-100.
- MacSwan, J., & Pray, L. (2005). Learning English bilingually: Age of onset of exposure and rate of acquisition of English among children in a bilingual education program. *Bilingual Research Journal*, 29(3), 687-712.
- MacSwan, J. & Rolstad, K. (2005). Modularity and the facilitation effect: Psychological mechanisms of transfer in bilingual students. *Hispanic Journal of the Behavioral Sciences*, 27(2), 224-243.
- MacSwan, J., & Rolstad, K. (2006). How Language Proficiency Tests Mislead Us About Ability: Implications for English Language Learner Placement in Special Education. *Teachers College Record*, 108, 11, 2304-2328.
- MacSwan, J., & Rolstad, K. (2010). The role of language in theories of academic failure for linguistic minorities, pp. 173-193. In Petrovic, J. E. (Ed.). *International perspectives on bilingual education: Policy, practice, and controversy*. Charlotte, N.C: IAP-Information Age Publishing, Inc.
- MacSwan, J., Rolstad, K., & Glass, G. (2002). Do some school-age children have no language? Some problems of construct validity in the Pre-LAS Español. *Bilingual Research Journal*, 26(2), 213-238.
- Mahoney, K. S., & MacSwan, J. (2005a). Re-examining identification and reclassification of English Language Learners: A critical discussion of select state practices. *Bilingual Research Journal*, 29(1), 31-42.
- Mahoney, K. S., & MacSwan, J. (2005b). *Identification, Reclassification, and Assessment of English Learners: A Survey of State Policies*. A monograph of the Language Policy Research Unit. Tempe, AZ: Arizona State University. Available at [http://www.terpconnect.umd.edu/~macswan/Mahoney %20and %20MacSwan % 202005 %20monograph.pdf](http://www.terpconnect.umd.edu/~macswan/Mahoney%20and%20MacSwan%202005%20monograph.pdf)
- Menken, K. (2009). «No Child Left Behind and its effects on language policy». *Annual Review of Applied Linguistics* 29, 2009, pp. 103-117.
- Messick, S. (1989). Validity, pp. 13-103. In R. L. Linn (Ed.), *Educational measurement*. New York: Macmillan.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23, 13-23.
- Milroy, J., & Milroy, L. (1999). *Authority in language: Investigating Standard English*. New York: Routledge.
- Newmeyer, F. (1986). *The politics of linguistics*. Chicago: The University of Chicago Press.
- No Child Left Behind [NCLB] (2001). *No Child Left Behind. Title III: Language instruction for limited English proficient and immigrant students*. 107th Congress, 1st Session,

- December 13, 2001. Washington, DC: George Washington University, National Clearinghouse for Bilingual Education.
- Pinker, S. (1994). *The language instinct: How the mind creates language*. New York: W. Morrow.
- Rolstad, K. (2005). Two-way bilingual instruction of third language children: Immersion or submersion? In B. Hufeisen & R. Fouser (Eds.), *Introductory Readings in L3*, pp. 83-100. Tübingen: Stauffenburg Verlag.
- Rolstad, K. (2014). Rethinking language in school. *International Multilingual Research Journal*, 8(1), 1-8.
- Rolstad, K., Mahoney, K., Glass, G. V. (2005). The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational Policy*, 19, 572-594.
- Ruiz, R. (1984). Orientations in language planning. *NABE Journal*, 8, 15-34.
- Schleppegrell, M. J. (2001). Linguistic features of the language of schooling. *Linguistics and Education*, 12, 4, 431-459.
- Slobin, D., & Bowerman, M. (1985). *Crosslinguistic evidence for the language making capacity: What shapes children's grammar?* Hillsdale, NJ: Erlbaum.
- Snow, C., & Uccelli, P. (2009). The challenge of academic language, pp. 112-133. In Olson, D. R., & N. Torrance (Eds.), *The Cambridge Handbook of Literacy*. Cambridge: Cambridge University Press.
- Sperber, D. & Wilson, D. (1986). *Relevance: Communication and cognition*. Cambridge: Cambridge University Press.
- Tager-Flusberg, H. (1997). Putting words together: Morphology and syntax in the preschool years. In J. Berko-Gleason (Ed.), *The development of language* (pp. 159-209). Boston: Allyn and Bacon.
- TESOL International Association. (2013). *Overview of the Common Core State Standards Initiatives for ELLs*. Alexandria, VA: Author.
- Wiley, T. (2005). *Literacy and Language Diversity in the United States*, Second Edition. Washington, DC & McHenry, IL: Center for Applied Linguistics & Delta Systems.
- Wiley, T. (2013). A Brief History and Assessment of Language Rights in the United States, pp. 61-90. In J. W. Tollefson (Ed.). *Language Policies in Education: Critical Issues, Second Edition*. London: Routledge.
- Wiley, T., & Rolstad, K. (2014). The Common Core State Standards and the Great Divide. *International Multilingual Research Journal*, 8(1), 38-55.
- Wong Fillmore, L. and Snow, C. (2002). *What teachers need to know about language*. McHenry, IL: Center for Applied Linguistics and Delta Systems.