

Reexamining Identification and Reclassification of English Language Learners: A Critical Discussion of Select State Practices

Kate S. Mahoney
Arizona State University–East Campus

Jeff MacSwan
Arizona State University–Tempe Campus

Abstract

In this article, the authors report select results of a national survey of state requirements and recommendations regarding identification and reclassification of English Language Learners (ELLs) conducted in academic year 2001–2002, called the *Survey of State Policies for Identification and Reclassification of Limited English Proficient Students*. The purpose of the *State Survey* was twofold: (1) to obtain data regarding current state practices with respect to identification and reclassification of ELLs; and (2) to raise questions regarding the appropriateness of three dominant practices, namely, (a) the use of academic achievement tests for the purpose of identification, (b) routine assessment of children’s oral native-language ability, and (c) the use of cutoff scores in determining identification or reclassification of ELL status. It is argued that such practices may lead to errors in identification and reclassification of ELLs, which in turn may have negative consequences for students.

Introduction

This article reports on select state procedures for reclassification and identification of English language learners (ELLs) and raises questions about the appropriateness of three dominant practices that we believe should be reconsidered, as they may lead to the incorrect identification or reclassification of ELLs. Such errors of identification or reclassification may have negative outcomes for students because they are associated with treatments developed

for children with different needs. To discover current practices, we conducted a national survey of state requirements and recommendations regarding identification and reclassification of ELLs conducted in academic year 2001–2002, called the *Survey of State Policies for Identification and Reclassification of Limited English Proficient Students* (henceforth, *State Survey*).

In the present paper, we critically evaluate (a) the use of academic achievement tests for the purpose of identification, (b) the routine assessment of children's oral native-language ability, and (c) the use of cutoff scores for identification or reclassification purposes. For those interested, the complete survey results, including the survey instrument and list of participants, are published separately as a monograph (Mahoney & MacSwan, 2005).

While we support bilingual education programs (August & Hakuta, 1998), we worry that some children may be poorly served as a result of misidentification and inappropriately late or early reclassification from language assistance. We believe that an understanding of current state practices and an informed critique of them may lead to improvement in our ability to meet the needs of language-minority children more effectively.

Literature Review

Although surveys of state practices have been conducted in the past, in addition to collecting typical information regarding the identification and reclassification process for ELLs, the *State Survey* reports current information on specific testing instruments used for both English and native-language assessment as well as information about the governing process of these assessments. The Council of Chief State School Officers (CCSSO) (1991) conducted a survey to encourage the standardization of procedures used to identify ELLs and to improve states' collection, reporting, and utilization of data about ELL students' educational status and progress. CCSSO's survey is similar to the *State Survey* discussed here in that it collects summary information about identification and reclassification methods, and includes specifics such as testing instruments and cutoff scores used by each state. In addition, Donley, Henderson, and Strand (1995) conducted a survey of a wide variety of issues regarding ELLs, including enrollment, academic progress, definitions and methods of identification, and enrollment in special language programs. Kindler (2002) conducted a survey that included demographics, program information, identification and reclassification methods, and testing instruments. Besides providing an update on current state policies regarding the identification and reclassification of ELLs, the *State Survey* reported here complements previous work of this type while providing more specific and detailed information on some specific questions—including, in particular,

official state practices regarding the assessment of children's native-language ability, and the names of the specific instruments used to assess students' first-language and second-language proficiency.

Method

In 2000–2001, we designed and implemented the *State Survey*, organized into three parts. The first part addressed the nature of the identification process for ELLs; the second part addressed the nature of the reclassification process; and the third part addressed specific testing instruments used for both identification and reclassification. Survey data were then aggregated and reported in a total of 21 tables and figures, available in Mahoney and MacSwan (2005).

Each appropriate official responsible for the oversight of programs for ELLs in the 50 U.S. states, the District of Columbia, and U.S. territories and outlying areas was contacted to participate in this survey. In all, 59 surveys were distributed. In 2001–2002, survey responses were received from all states except Indiana, Iowa, Montana, and Washington, as well as from the District of Columbia, the Virgin Islands, and Puerto Rico, for a response rate of 83%. Below, for expository convenience, we use the term “states” to refer collectively to all federal entities surveyed.

Results

In this short discussion, we focus on three specific issues of particular concern to us: (a) the use of academic achievement tests for identification purposes, (b) the practice of routinely assessing children's oral native language ability, and (c) the use of cutoff scores for either identification or reclassification.

The Inappropriate Use of Academic Achievement Tests for Identification

Although they function closely together, language proficiency and academic achievement are two distinct constructs and should be measured separately (MacSwan & Rolstad, 2003; Thompson, Dicerbo, Mahoney, & MacSwan, 2002). Indeed, there is substantial evidence that knowledge of language and general knowledge (including knowledge of school subject matter) are represented in distinct components of the mind or brain (MacSwan & Rolstad, 2005). It is widely known that all (typically developing) children acquire the language of their speech community, while they differ a great deal with respect to their success in school.

Cross-cultural research in language acquisition has found that all children acquire the language of their respective speech communities, and do so effortlessly and without instruction (Slobin & Bowerman, 1985; Pinker, 1994).

During the most active acquisition period in the preschool years, children learn approximately 10–12 new words per day, frequently on one exposure and under very ambiguous circumstances (Gleitman & Landau, 1994). As Tager-Flusberg (1997, p. 188) has noted, “by the time children begin school, they have acquired most of the morphological and syntactic rules of their language,” and possess a grammar essentially indistinguishable from that of adults. These facts and others have led researchers to believe that language acquisition is directed by innate principles of mental architecture (Chomsky, 1986).

Literacy and other school subjects certainly make use of a child’s language ability, but these seem substantially different in character. Humans acquire language by instinct, 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, 2000; MacSwan & Rolstad, 2003, 2005). Academic achievement denotes a domain of knowledge that is specific to a particular human context — namely, the world of formal schooling. While all children develop a vocabulary specific to their interests and endeavors and a grammatical system consistent with the language of their speech community, not all children will come to know that the square of the hypotenuse of a right-angled triangle is equal to the sum of the squares of the other two sides.

Whether and to what extent second language learning is more like first language acquisition or more like learning to play chess, or something in-between, is an open question (Bialystok & Hakuta, 1994). But it is certain that children might not be proficient in English and yet be perfectly competent at academic tasks, and vice versa. Hence, the use of English-medium academic achievement tests to identify children as ELL students is very problematic. Since such tests are not constructed with reference to a theory of language ability, they cannot be used as proxy measures of language proficiency. A low score might reflect limited knowledge of school subjects or limited proficiency in the language of the test. Thus, the practice of using such tests as a component of the identification process is a mistake, in our view, as students’ limited proficiency in the language of the test becomes a source of error in measures of academic achievement.

The Inappropriate Practice of Routinely Assessing Children’s Oral Native-Language Ability

Figure 1 indicates that 14 states (about 29% of respondents) report using oral native-language assessments as one method of identifying ELL students. The number of states using this method increased from 9 to 14 in the 11 years between the CCSSO survey, discussed above, and the *State Survey*. Furthermore, as shown in Figure 2, our survey found that the most popular primary-language assessments were the Language Assessment Scale–Español

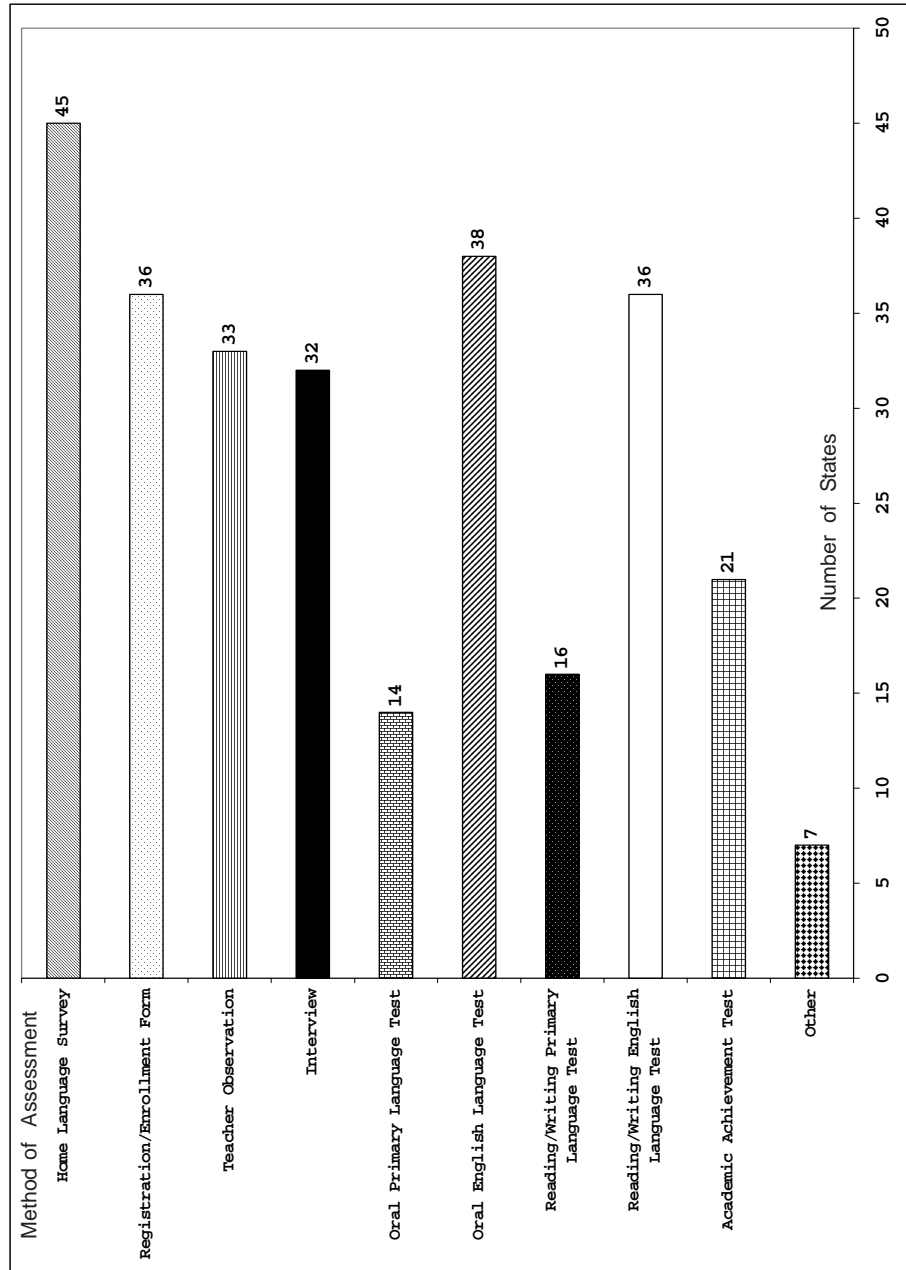


Figure 1. Methods used by states to identify ELL students.

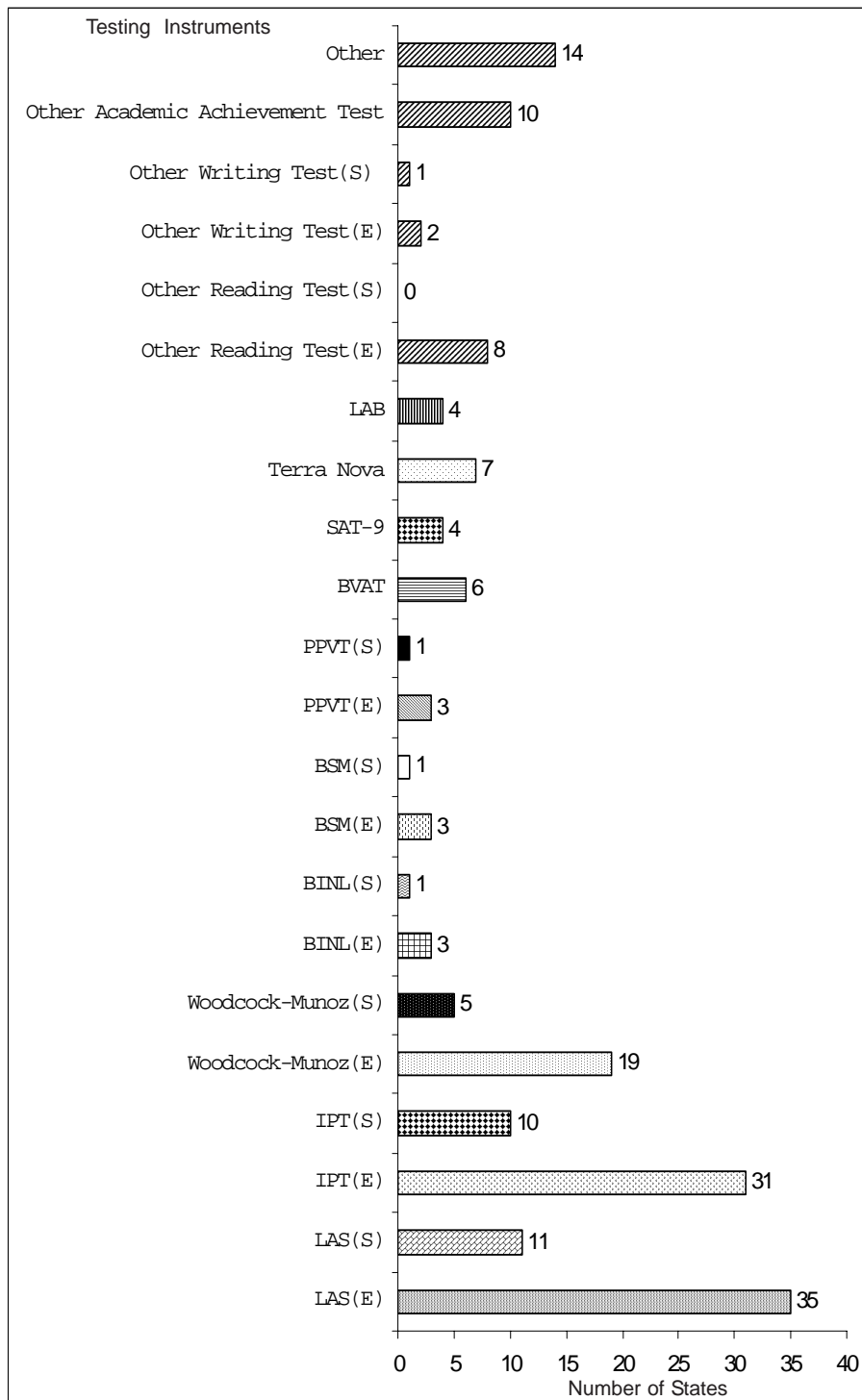


Figure 2. Testing instruments used by states for identification and reclassification.

(LAS[S]) (11 states), the IDEA Language Proficiency Test–Spanish (IPT[S]) (10 states), and the Woodcock-Muñoz Language Survey (Woodcock-Muñoz [S]) (5 states).

Given observations made earlier regarding the nature of language and acquisition, one would not expect children’s oral native language to require assessment upon entry to school simply because they are linguistic minorities. While it may be useful for schools to assess children’s reading and writing ability in their home language upon entering school, using potentially misleading native-language assessments adds confusion rather than additional clarity to the identification process.

For example, in a study of the Pre-Language Assessment Scales–Español (Pre-LAS Español), MacSwan, Rolstad, and Glass (2002) assessed the validity of this version of the LAS–Español, which purports to measure the oral native–language ability of Spanish-speaking children ages 4 to 6. The authors examined a dataset of 38,887 students who took the Pre-LAS Español in a large urban school district in 1997. Approximately one third of children who were identified as originating from Spanish-speaking homes were classified by the test as “non-” or “limited” speakers of their own language by the test. It was found that the final section of the test, a story-retelling task, accounted for 60% of the variance in total scores (correcting for overlap). Furthermore, the data revealed that 67.5% of “non-Spanish speakers” scored 80% or higher on most of the test, and that 20% of the “non-Spanish speakers” who scored 80% or higher on the first four subparts of the test had not given any response to the story-retelling section at all. Hence, rather than detecting children’s level of proficiency in Spanish, it appears to have rated their readiness to engage in the retelling task, possibly affected by their perception of the task as inane and outside the scope of meaningful uses of human language. The authors provide a detailed critique of the portions of the test most responsible for the high proportion of non- and limited Spanish scores.

Similarly, MacSwan and Rolstad (2003) note problems with the Spanish version of the Idea Proficiency Test (IPT), which erroneously lead to flawed assessment of language minority children. To do well on the test, children are required to answer certain items in complete sentences, recalling Bereiter, Engelman, Osborn, and Reidford’s (1966) method of assessing African American Vernacular English, noted and classically critiqued in Labov (1970). Making matters all the worse, the developers of the IPT–S knew so little about language that they construed a “complete sentence” in Spanish as requiring an overt subject, as required by prescriptivists in the *English*-speaking world.

School administrators are often frustrated and confused by test results indicating that an ELL lacks proficiency in both English and his or her native language. Reasonable questions arise about what language of instruction is appropriate for children classified as “non-proficient” in Spanish, even though they come from Spanish-background homes. In addition, Artiles, Rueda,

Salazar, and Higareda (2005) report that ELL children assessed as lacking proficiency in their native language have a high likelihood of being classified (arguably incorrectly) as special education students. Although it has been argued that assessing children's native language provides supplemental information to help teachers and administrators better evaluate students' English-proficiency test results (CCSSO, 1991), we believe it is more likely to create an atmosphere of confusion and result in incorrect perceptions of children's learning situations.

It is doubtful that erroneous results such as these can be repaired by correcting the problems noted. Rather, we believe that the problems follow from a fundamental incompatibility between the assessment purpose and classical methods of test validation. Because knowing a native language is more like knowing how to walk than like knowing mathematics or history, properly designed tests of native-language ability, administered on a random sample of native speakers, will always obtain dramatic ceiling effects. Such results lead test developers to regard tests as either incorrectly constructed (the mistaken conclusion) or administered to the wrong population (the correct conclusion).

We therefore recommend that states abandon routinely assessing children's native-language ability. However, in some instances, a Spanish-speaking child, like any other, may be suspected of having a language-related learning disability, and should be referred for assessment to determine whether the child might need special education services. In addition, we make a strong distinction between assessing a child's native-language ability and assessing a child's academic subject matter knowledge in his or her native language. The latter, like the assessment of children's knowledge of reading and writing in their native language, improves our understanding about the role that prior academic experience in the home language might play in students' ongoing educational experience. The former does not.

The Use of Arbitrary Cutoff Scores

The *State Survey* revealed that 15 states used a specific cutoff score on a test of academic achievement, reading, or writing for identification purposes, and another 15 used a specific cutoff score on such tests for reclassification purposes (Mahoney & MacSwan, 2005).

Complications surrounding the use of cutoff scores have been evident since the 1970s, when criterion-referenced testing gained popularity (Bracey, 2002). One area of controversy is the use of one specific score as a point of determining high-stakes decisions such as graduation and grade promotion. The measurement community does not condone the use of one specific score as a threshold for making high-stakes decisions, because doing so ignores the Standard Error of Measurement (SEM). Each published standardized test

has an associated SEM, given in its technical manual. A SEM of 4 is an indication that the score is not perfectly reliable and that the true score lies somewhere within the range of plus or minus 4 points of the scaled score. For example, if a student scores 64 on a test and the SEM = 4, the test publisher is not confident that the true score is 64, but rather maintains that the child's true score or true ability lies somewhere between 60 and 68. If, say, a cutoff score is given as 64, then a student scoring 61 would be within the range by virtue of SEM, but would be erroneously excluded as a matter of policy. Bracey (2002) notes that these important facts are virtually ignored by policymakers who attach high-stakes decisions to specific cutoff scores on standardized tests. In addition, as discussed in Glass (1978), a question arises about defining a point where incompetency ends and competency begins. Glass argues that the line drawn is necessarily arbitrary, determined by politics rather than principle.

From a validity perspective, what matters most is not the validity of the test, but validity of test *use*. Each distinct use of a test must be scrutinized, and evidence must be available in support of it (Messick, 1990). Thus, if a state chooses to use the 40th percentile to determine whether an ELL should be reclassified as a non-ELL student and placed into a mainstream classroom, then empirical evidence must be made available to validate whether or not scoring at the 40th percentile on a test is a good predictor of success in mainstream classrooms. Unless and until such evidence is available, cutoff scores should not be imposed by policymakers.

Conclusions

We have discussed selective results from the *Survey of State Policies for Identification and Reclassification of Limited English Proficient Students*, presented in full in Mahoney and MacSwan (2005). We have been critical of three common practices reported to be fairly widely used, including (a) the use of academic achievement tests for the purpose of identification, (b) the practice of routinely assessing children's oral native-language ability (but not of assessing literacy and content-area knowledge *in* the native language), and (c) the use of cutoff scores for either identification or reclassification purposes. We invite others to join in a critical evaluation of these and other practices presented in our full report (Mahoney & MacSwan, 2005).

A limitation of the survey is that it was conducted immediately before a sweeping overhaul of federal education policy was signed into law as the No Child Left Behind Act (NCLB) (2002). Concurrently with the passage of NCLB, the Bilingual Education Act of 1968, which had been repeatedly reauthorized, was repealed and replaced with the English Language Acquisition Act, which requires states to develop a single language testing instrument for all ELL students. (See Crawford, 2004, for discussion.)

While general policies regarding identification and reclassification will, as always, remain in flux, the specific testing instruments used by states to assess ELLs' language proficiency reported here will no doubt begin to change as states yield to the requirements of NCLB. While this is a limitation of the data reported here, it is also a strength, as it provides a snapshot of specific testing policies in place immediately before the passage of NCLB.

We advocate a child-study approach to assessment of ELL students, one that takes into account a wide range of evidence bearing on an individual child's specific needs and in which all stakeholders have a voice in important decisions. Local resources and program options are as important as the child's level of proficiency in the second language, and must also be taken into consideration. Criteria for identification might be rather different from those established for reclassification, and in no case should important decisions be made based on one or more scores on standardized tests of language ability or academic achievement.

References

- Artiles, A. J., Rueda, R., Salazar, J. J., & Higaeda, I. (2005). Within-group diversity in minority disproportionate representation: English language learners in urban school districts. *Exceptional Children, 71*(3), 283–300.
- August, D., & Hakuta, K. (Eds.). (1998). *Educating language minority children*. Washington, DC: National Academy Press.
- Bereiter, C., Engelman, S., Osborn, J., & Reidford, P. A. (1966). An academically oriented pre-school for culturally disadvantaged children. In F. M. Hechinger (Ed.), *Pre-school education today* (pp. 105–135). New York: Doubleday.
- Bialystok, E., & Hakuta, K. (1994). *In other words: The science and psychology of second language acquisition*. New York: Basic Books.
- Bracey, G. (2002). *Put to the test: an educator's and consumer's guide to standardized testing*. Rev. ed. Bloomington, IN: Phi Delta Kappa International.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.
- Council of Chief State School Officers (CCSSO). (1991). *Summary of state practices concerning the assessment of and the data collection about limited English proficient (LEP) students*. Washington, DC: CCSSO, State Education Assessment Center/Resource Center on Educational Equity.
- Crawford, J. (2004). *Educating English learners: Language diversity in the classroom* (5th ed.). Los Angeles: Bilingual Education Services.

- Donley, B., Henderson, A., & Strand, W. (1995). *Special Issues Analysis Center. Annual report: Year 3. Vol. 3: SEA Report, Task 7. Summary of state educational agency program survey of states' limited English proficient persons and available educational services, 1993–1994.* (ERIC Document Reproduction Service No. ED389179)
- Gee, J. P. (2001). Progressivism, critique, and socially situated minds. In C. Dudley-Marling & C. Edelsky (Eds.), *The fate of progressive language policies and practices* (pp.31–58). Urbana: National Council of Teachers of English.
- Glass, G. V. (1978). Standards and criteria. *Journal of Educational Measurement, 15*, 237–261.
- Gleitman, L., & Landau, B. (1994). *The acquisition of the lexicon.* Cambridge, MA: MIT Press.
- Kindler, A. (2002). *Survey of the states' limited English proficient students and available educational programs and services 2000–2001, summary report.* Washington, DC: Prepared for Office of English Language Acquisition by the National Clearinghouse for English Language Acquisition & Language Instruction Educational Programs. Retrieved March 18, 2005, from <http://www.ncele.gwu.edu/policy/states/reports/seareports/0001/sea0001.pdf>
- Labov, W. (1970). The logic of non-standard English. In F. Williams (Ed.), *Language and poverty* (pp. 225–261). Chicago: Rand McNally.
- MacSwan, J. (2000). The Threshold Hypothesis, semilingualism, and other contributions to a deficit view of linguistic minorities. *Hispanic Journal of Behavioral Sciences, 22*(1), 3–45.
- MacSwan, J., & Rolstad, K. (2003). Linguistic diversity, schooling, and social class: Rethinking our conception of language proficiency in language minority education. In C. B. Paulston & R. Tucker (Eds.), *Sociolinguistics: The essential readings*, (pp. 329–340). Oxford, England: Blackwell.
- MacSwan, J., & Rolstad, K. (2005). Modularity and the facilitation effect: Psychological mechanisms of transfer in bilingual students. *Hispanic Journal of Behavioral Sciences*, forthcoming.
- MacSwan, J., Rolstad, K., & Glass, G. V. (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. (2005). *Identification, reclassification, and assessment of English learners: A survey of state policies.* Tempe: Arizona State University. Retrieved March 30, 2005, from <http://www.asu.edu/educ/eps1/LPRU/features/>

- Messick, S. (1990). Validity of test interpretation and use. ETS-RR-90-11. Princeton, NJ: Educational Testing Service.
- No Child Left Behind Act, Pub. L. No. 107-110 (2002).
- Pinker, S. (1994). *The language instinct: How the mind creates languages*. New York: William Morrow.
- Slobin, D., & Bowerman, M. (1985). *Crosslinguistic evidence for the language making capacity: What shapes children's grammar?* Hillsdale, NJ: Erlbaum.
- 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 & Bacon.
- Thompson, M., Dicerbo, K., Mahoney, K., & MacSwan, J. (2002). Exit to California? A validity critique of language program evaluations and analysis of English learner test scores. *Education Policy Analysis Archives*, 10(7). Retrieved March 18, 2005, from <http://epaa.asu.edu/epaa/v10n7/>