Different Questions, Different Answers

A Critique of the Hakuta, Butler, and Witt Report, “How Long Does It Take English Learners To Attain Proficiency?”

Christine H. Rossell, Ph.D.

Kenji Hakuta, Yuko Goto Butler, and Daria Witt begin their paper with the statement:

One of the most commonly asked questions about the education of language minority students is how long they need special education services, such as English as a Second-Language (ESL) and bilingual education (p.1).

Unfortunately, they do not present any research on this issue in their paper. Nevertheless, this does not stop them from concluding:

The data would suggest that policies that assume rapid acquisition of English—the extreme case being Proposition 227 that explicitly calls for “sheltered English immersion during a temporary transition period not normally intended to exceed one year”—are wildly unrealistic (p. 13).

Although they appear not to know it, there is no research presented in this paper that tells us how long limited-English proficient (LEP) students should be in a sheltered English immersion classroom. The research that is presented is on a different issue: how long it takes a limited-English proficient student, on average, to attain the average English language achievement of fluent English speakers or a test publisher’s criterion for English proficiency.

The authors are simply wrong in believing that knowing how long it takes an LEP child to achieve parity with native English speakers, or to be classified “proficient” on an English proficiency test, tells us how long they need special education services or how long they should be in a sheltered immersion classroom.

The Data

The Hakuta et al. study consists of LEP students in four samples, two of them in school districts in the San Francisco Bay area and two of them in Canada. They collected and analyzed the data in School Districts A and B in California themselves and reanalyzed summary data on the two Canadian samples that were reported in Wright and Ramsey, 1970; Cummins, 1981; and Klesmer, 1993.

School districts A and B in California vary considerably in socioeconomic status (SES). The sample of LEP students in district A consists of all 1,872 LEP students in Grades 1-6 in spring 1998 who had been in the district since kindergarten and were classified at that time as LEP. About half were Vietnamese speakers and half Spanish speakers. According to the authors, the district has been on a state waiver from bilingual education, and has never provided systematic instruction through the native language. The percentage of students on free or reduced-price lunch is low—35 percent—and their annual redesignation rates from LEP to English proficient are high, about four times the state average.

District B, by contrast, has a free or reduced lunch rate of 74 percent—twice that of District A. The sample in District B consists of 122 Spanish speakers in grades 1, 3, and 5 during the spring of 1998, randomly selected from the students who had been in the school district since kindergarten, were classified LEP at that time, and who attended high poverty schools. Some of these LEP students were in bilingual education and some in ESL, although the authors assert there was no difference in achievement between students in the two programs.

The Toronto data reported in Wright and Ramsey (1970) and Cummins (1981) consists of 1,200 immigrant children learning English as a second language selected from a survey of 25 percent of the Toronto school system’s classrooms in Grades 5, 7, and 9, who were of varying length of residence in Canada. Although the authors do not specify what language the students were instructed in, it was undoubtedly English since that is the normal approach in Canada to educating immigrant children.

The North York, Ontario data, reported in Klesmer (1993), consisted of a randomly selected sample of 285 ESL students and 43 native English-speaking students who were controls. All students were 12 years old and most of them were in the seventh grade, but their length of residence ranged from six months to almost six years. Since the students are called “ESL” students, we can assume they are being instructed in English.

The research design varies across studies. The data from Toronto and North York are cross-sectional. They consist of students at fixed grade levels who differ in their length of residence in Canada. The data from Districts A and B in California are longitudinal and consist of the more sta-
able LEP students, those who had been in the school district since kindergarden and were classified LEP at that time. The Canadian data are not longitudinal, but they will be biased only if the composition of the students being studied changes over time in a way that influences the outcome. I am not aware of any such changes, and the authors do not mention any.

Hakuta, Butler and Witt’s findings are divided into oral English and academic English, a distinction that is commonly made, but in fact is not based on research or experience. All English is academic English, and there really is no way to separate academic English from oral English. Moreover, there is extensive research, discussed below, that contradicts the notion that oral English proficiency is “non-academic” and only tests whether a child understands the English language. Therefore, I have changed their term “academic” to “written” to conform to what the tests actually assess and to maintain the useful distinction between oral and written tests.

Hakuta, Butler, and Witt’s findings on how long it takes LEP students to attain English “proficiency” are summarized in Table 1. English proficiency was assessed in the two California school districts by means of a specific criterion on oral and written English proficiency tests and in the two Canadian samples by means of parity with native English speakers on oral and written standardized achievement tests. Their findings show that, on average, it takes anywhere from two years to perhaps forever to attain the criterion for English proficiency in the California school districts, and from nine years to perhaps forever to attain parity with English native speakers in the Canadian samples.

Table 1.
Number of Years It Takes LEP Students To Attain English Language Parity with Native English Speakers or a “Proficient” Score on an English Proficiency Test in the Hakuta, Butler, and Witt Report

<table>
<thead>
<tr>
<th></th>
<th>CALIFORNIA</th>
<th>CANADA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ENGLISH PROFICIENCY TESTS</td>
<td>STANDARDIZED ACHIEVEMENT TESTS</td>
</tr>
<tr>
<td>(criterion)</td>
<td>(parity)</td>
<td>(parity)</td>
</tr>
<tr>
<td>DISTRICT A</td>
<td>DISTRICT B</td>
<td>TORONTO</td>
</tr>
<tr>
<td></td>
<td>(higher SES)</td>
<td>(lower SES)</td>
</tr>
<tr>
<td>ORAL</td>
<td>2 - 5 years</td>
<td>9 - 11 years</td>
</tr>
<tr>
<td>WRITTEN</td>
<td>4 - 7 years after 5 years</td>
<td>Not attained after 5 years</td>
</tr>
</tbody>
</table>

Unwarranted Conclusions

Hakuta, Butler, and Witt jump to the conclusion that the number of years it takes LEP students to reach the average for native English speakers or the publisher’s criterion for English proficiency is the number of years they need special education services. There are two reasons why this conclusion is unwarranted. First, parity with English speakers on English proficiency tests or standardized achievement tests is a badly flawed standard for determining fluency in English. Half of all native English speakers cannot achieve the average standardized test score for native English speakers, and almost as large a percentage cannot achieve the publisher’s criterion for English proficiency. If the students are of low SES, as is typically the case with immigrant children, more than half will not achieve the average for English speakers or the criterion for English proficiency, no matter how fluent they are in English. This failure to understand what test scores mean and their biases is probably one of the most common errors made by reporters, politicians, other laymen, and even by experts in the field, and it is disheartening to see the mistake made once again.

The second reason why one cannot jump to the conclusion that the number of years it takes LEP students to reach the average for native English speakers or the publisher’s criterion for English proficiency is the number of years they need special education services is that the research design used by Hakuta et al., and the studies they analyze, do not allow us to draw such conclusions. To determine whether an LEP child is better off with special education services than without requires the following design. LEP children must be randomly assigned to a group that receives no special education services and to groups that receive some carefully documented special service over different periods of time. The achievement of students in these
groups is then compared and a statistical analysis performed to determine if there is a significant difference between the groups. The research design that would definitively answer the question of how long LEP students should remain in a structured immersion classroom, a particular type of special education service, would randomly assign non-English speaking students in each grade to a mainstream classroom and to a structured immersion classroom. These students would be tested initially and then at monthly intervals. The point at which the students in the mainstream classroom outperform the students in the structured immersion classroom on the tests is the point at which a student is better off in the mainstream classroom than in a structured immersion classroom. If over several years, the students who entered the mainstream classroom sooner outperform the students who entered later, then a mainstream classroom is a better environment from the start regardless of the short-term data.

Although it might appear that the structured immersion classroom would be superior to the mainstream classroom for a very long time, that is probably not the case. My own estimate would be that sometime during the first year there is probably no difference between the structured immersion classroom and the mainstream classroom because although the structured immersion classroom may be a better environment in the beginning, it has the following negative characteristics: (a) it has a slower pace which will begin to negatively affect students who can understand English, and (b) it has no English-speaking role models. Students interact with other students whose English is also imperfect and this can become a problem because students emulate the English of their peers. If they cannot understand English, they are better off in a structured immersion classroom. But when they reach the point where they can understand English, they will speak like their classmates and they will be better off if their classmates are speaking grammatical English.

Hakuta et al. do not do this analysis, nor do they present research that has done this. Therefore, they cannot legitimately claim that their study tells us how long LEP children should receive special education services or be in a structured immersion classroom.

Norm-Referenced Tests
As noted above, the first reason why the findings of Hakuta et al. cannot tell us how long a child needs special education services is that the instruments and procedures used to measure English proficiency are flawed both in design and in use. Table 2 summarizes the tests and standards used in the samples analyzed by Hakuta, Butler, and Witt. The table is divided into the same categories as Table 1, but the cells now contain the type of test and the criterion used for English proficiency. In addition, I have added a row indicating the biases of the English proficiency tests used in the California school districts and the standardized achievement tests used in the Canadian samples in determining whether a student is fluent in English.

English proficiency tests are a type of norm-referenced test given to students identified by a home language survey as coming from a home where someone speaks, or has spoken, a language other than English. They are also given to students who have already been identified as LEP in order to determine if they are now English proficient. The state of California approves the following English proficiency tests which have both oral and written forms: the BINL, BSM I/II, Pre-IPT, IPT I/II, pre-LAS, LAS I/II, the Woodcock-Muñoz Language Survey, and the QSE.

Two of these tests are used in the California districts studied by Hakuta, Butler, and Witt. District A uses the IPT and the oral results are shown in Table 1. District B uses the Woodcock-Muñoz Language Survey. The results for the written portion are also in Table 1.

As shown in Table 2, the California school districts use a specific English proficiency criterion established by the publisher in the case of the IPT and the Woodcock-Muñoz Language Battery and by the district in the case of the MacMillan Informal Reading Inventory. The Canadian studies use parity with English speakers on oral and written standardized achievement tests.

Children can be completely fluent in English, indeed they can know no language other than English, and yet fail to achieve the publisher's criterion for English proficiency. All language proficiency tests, whether they are administered only to LEP students (and called English proficiency tests) or to English speaking students (and called achievement tests), are norm-referenced on fluent English speakers and are tests of the ability to speak and understand a language and tests of academic ability in that language. The publishers select a score on the English proficiency tests that they claim denotes whether a student is a fluent English speaker, but in fact there are English monolingual students who will score below whatever score they select unless it is zero. Typically the publishers select a score that can only be achieved by about 60 percent to 70 percent of the English monolingual students.
### Table 2.
Tests and Standards Used in Hakuta, Butler, and Witt Report and Their Biases

![Table 2 Image](attachment:image.png)

Using parity with native speakers on standardized achievement tests as a means of determining English fluency, as is done in the Canadian studies, is biased by the fact that standardized achievement tests rank order students and this rank ordering is highly correlated with socioeconomic status. The test scores do not tell us what students know. They only tell us who knows more and who knows fewer answers to the items on the test. These items are deliberately selected to produce a normal curve among English speaking students, and the test scores are highly correlated with socioeconomic status.

The analyses of the Canadian samples presented in Hakuta, Butler, and Witt are biased by the fact that immigrant children are of lower social class than non-immigrant children. This is shown in Figure 1 which presents data on the percentage of students on free or reduced lunch by LEP status in spring 1997 in a medium sized California school district of about 35,000 students. The percentage of currently LEP students who are poor is 71 percent and the percentage of currently or formerly LEP students who are poor is 65 percent. The latter group includes formerly LEP students so as to include as many immigrant students as possible, not just those who continue to score low on English proficiency and standardized achievement tests. LEP students have more than three times the percentage poor of non-LEP students.

![Figure 1](attachment:image.png)

To understand how this affects the standardized achievement test results in the Canadian samples in Hakuta, Butler, and Witt, we need to look at the relationship between poverty and standardized test scores in an English speaking sample. Figure 2 shows a box and whiskers plot of the CAT5 achievement—vocabulary, reading comprehension, math analysis, and math computation—of all secondary students (including poor students) at the top of the page and the achievement of only the poor students under it,
among students who are fluent English speaking and who have never been classified LEP. The black line across each box is the median achievement for each group. The box itself is the interquartile range—the range from the 25th to the 75th percentile that contains 50 percent of the cases. The horizontal lines at each end of the vertical lines are the maximum scores.

I have added the average scores for each group of students below each subtest. Note the 36th percentile, the most common standard in California for redesignating an LEP student as English proficient.

The analysis in Figure 2 is similar to the analysis of the Canadian studies presented in Hakuta, Butler, and Witt. The studies they analyzed compared the standardized test scores of immigrant children to English native speakers. To show the bias produced by the fact that immigrant children are of lower social class than all children, I have compared poor English-speaking children to all English native speakers. English speaking poor children are not even close to attaining parity with all English speakers, although both groups are fluent English speakers. Indeed, this is exactly the problem with standardized, achievement tests—they merely rank order students on knowledge of the items on a test and they cannot tell the difference between students who do not know English and students who do not know the answer. Thus, any comparison of the achievement of a high-poverty group, such as immigrant children, to the achievement of all students, as Hakuta, Butler, and Witt have done, will find the poorer group performs worse.

**English Proficiency Tests**

All English proficiency tests, whether oral or written, are known to be unreliable—that is, you cannot get the same outcome in subsequent tests of the same child—and invalid—that is, they do not accurately determine who is LEP (Baker and Rosell, 1987; Rosell and Baker, 1988). On the face of it, oral English proficiency tests would seem to be better than a written test at determining whether a child knows enough English to function in a mainstream classroom because the child doesn’t have to know how to read or write to take an oral proficiency test.

Unfortunately, oral English proficiency tests are no better than written English proficiency and standardized achievement tests, and for many of the same reasons. Moreover, they have some additional problems that written proficiency tests do not have. In oral tests, students are asked questions that require that they not only know English, but understand and remember the question and have the self-confidence to stand up to a stranger when the question is not understood. Thus, contrary to the assertions of Hakuta, Butler, and Witt, oral tests are as "academic" as written tests. Like standardized achievement tests administered to the English speaking stu-
dent body, and written English proficiency tests administered only to LEP students, oral proficiency tests cannot tell the difference between a student who does not know English and a student who does not know the answer. They are normed on an English-speaking body and the same arbitrary cutoff points are used.

Despite these problems, language proficiency tests are used everywhere as a means of identifying whether a student is LEP and English proficient, and their use is codified in state legislation and court decisions. New York City, for example, uses the L.A.B. whose oral portion was normed in 1981-82 and whose written portion was normed in 1985, in both instances on an English-speaking citywide population. The criterion selected for determining whether a child is fluent English proficient in New York City is currently the 40th percentile on the L.A.B. In many California school districts, including the two studied by Hakuta, et al., the standard is the 36th percentile. It is a mathematical principle that 40 percent of the norming population scores at the 40th percentile, and 36 percent scores at the 36th percentile. If the L.A.B. were administered citywide in New York City, a minimum of 40 percent of the children in the city, almost all of whom are English native speakers, would fail to be classified as English proficient. If the tests used in the California districts were administered to all students, a minimum of 36 percent would fail to be classified as English proficient, even if the only language they know is English.

To the extent that these students are of lower SES, even higher percentages will fail to be classified as English proficient. If we look at the box on the bottom of Figure 2, the analysis of the achievement of poor, English speaking students, we can see that about 50 percent of these students would fail to be classified English proficient if the standard were the 36th percentile.

Interestingly, the average human being seems to prefer a standard that he or she knows is wrong to no standard at all. After listening to the conflicting testimony on English proficiency, the judge in Aspia of New York, Inc., et al. v. Board of Education of the City of New York, et al. (394 F. Supp. 1975) concluded:

The most vivid point to emerge from all the argumentation is that we confront an enormous amount of speculation and uncertainty... (Aspia, 1975:1161).

"Without approaching confidence or certainty," (p. 1164) the Court defined the plaintiff class as Hispanic students who scored at or below the 20th percentile on the English L.A.B., but higher on the Spanish L.A.B. The Court then went on to say:

The crudity of this formulation is acknowledged on all sides. It is not possible to say with precise and certain meaning that an English-version score at a given percentile is similar to the same percentile score on the Spanish version...But we are merely a court, consigned to the drawing of lines, and we do the best we can (p. 1168).

Not long after the 1975 Aspia decision, the National Institute of Education analyzed the whole area of relative language assessment for the U.S. Department of Education and found no agreement as to what language proficiency is and general agreement that language proficiency tests are unreliable and invalid.

...In addition to such problems as low reliability and questionable validity and variation in theoretical underpinnings, differences in quality and quantity of items selected, and the plain fact of the incredible complexity of language, there are serious practical problems associated with assessing language proficiency on the basis of these instruments. Recent empirical studies indicate that the placement of children varies (often significantly) depending on which test is used (Spolsky, in NIE, 1981:38).

More recently, Irujo, Kramsch, Dube and Yedlin (1986) surveyed the issue of language proficiency for the Massachusetts Department of Equal Educational Opportunity. They found over 20 different definitions and concluded that language proficiency is one of the most poorly defined concepts in the field of language education. Yet, Massachusetts school districts continue to use language proficiency tests to classify students as LEP or English proficient.

The IPT, used in District A of the Hakuta, Butler, Witt study, has been found to be quite unreliable by Ramirez, Yuen and Ramey (1986). Of 573 kindergarten students classified as non-English-speaking, limited-English-speaking or fluent-English-speaking in the fall of 1984 in California, 236 had moved up one category, 238 had stayed the same, and 99 had moved down one category or more two years later in the spring of 1986. Thus, according to this test, not only has 40 percent of the sample made no progress in English over two years, but 17 percent know less English than when they began.

Similar results are found with students in higher grades. Of 232 first-graders classified LEP by the IPT in the fall of 1984, 50 percent made no progress over two years, and 13 percent knew less English than when they began. Of 123 third-graders classified LEP, 48 percent seemingly made no progress and 7 percent knew less English than when they began (Ramirez, Yuen, and Ramey, 1986).

LEP students who score low in English often score low in their native tongue because the tests also measure academic ability, not just fluency. Illustrative of this phenomenon is a study of relative language proficiency among Hispanic students in California by Duncan and De Avila (1979). A majority (54) of the 101 students classified by the Language Assessment Scales (LAS) as limited or non-proficient in English were also classified as...
limited or non-proficient in Spanish. Of the 96 students found to be limited or non-proficient in English, less than half (42) were considered proficient Spanish speakers according to their Spanish test score.

Moreover, language proficiency tests do not agree with each other even when they are in the same language. Ulibarri, Spencer and Rivas (1980) investigating the comparability of three oral English proficiency tests used in California (the LAS, BSM, and BINL) concluded that language classification is a function of the particular test used with each test identifying different numbers of eligible students. Studies by Gillmore and Dickerson (1979), Cervantes (1982) and Pelavin and Baker (1987) found similar results. Pelavin and Baker further found that the disagreement between tests is greatest for those students who spoke some English, in particular when a reclassification decision was being made.

Not only are the tests unreliable, but they are invalid. English proficiency tests administered to English monolingual children in experiments routinely classify large percentages of them as LEP. Berdan et al. (1982) administered the Language Measurement and Assessment Instrument (LM&AI) to Cherokee students at the request of the Cherokee Nation, which wanted to determine the need for Cherokee bilingual education. Through home interviews, Berdan et al. found that 82 percent of the Cherokee students were English monolinguals. The LM&AI, however, classified 48 percent of these monolingual English-speaking children as LEP presumably in need of instruction in Cherokee so they could improve their English. In 1984, the U.S. Department of Education had the LM&AI administered to a nationally representative sample of monolingual English-speaking school-aged children. The test classified 42 percent of them as LEP (U.S. Bureau of the Census Data, 1984).

A similar experiment in Chicago (Perlman and Rice, 1979) suggests that the problem of classifying English monolingual students as limited-English-proficient is not limited solely to low-achieving students. The Chicago Board of Education administered the Language Assessment Scales (LAS)—a test used widely throughout the U.S. and one of the approved tests in California—to students who spoke only English and were above the citywide TTBS norms in reading. Almost half of the monolingual, above average, English-speaking children were misclassified as non- or limited-English speaking. Moreover, there is a developmental trend. Seventy-eight percent of the English monolingual 5-year-olds, but only 25 percent of the 14-year-olds, were classified LEP.

I am also familiar with a particular instance of misclassification in California. In 1988, the principal of an elementary school in the Berkeley Unified School District, upset over the State Department of Education's compliance review, decided not to wait for the results of the home language survey before testing all new Spanish-surnamed students in her school. The 5-year-old child of a professional Hispanic family in Berkeley was administered the oral portion of the IPT in this mass testing. Although he knew no language other than English and the language of their home is English, he failed the oral proficiency test, was classified as limited-English-proficient, and assigned to the Spanish bilingual program. When the family received the notice, the mother called the school, informed it of their mistake, and was allowed to withdraw her child from the bilingual education program. But what if the mother had not been a fluent English speaker and an assertive professional who understood the mistake? There is a very good chance that this child would have been assigned to the Spanish bilingual program and taught in a language he did not know. A year later this same child, who at age 5 had been classified LEP by an oral proficiency test, was classified "gifted" on the basis of a standardized achievement test. Thus, it is possible for a gifted child to fail an oral English proficiency test and be classified LEP!

To summarize, the research evidence indicates that language proficiency tests are unreliable and invalid and there is a good deal of disagreement between the different types, particularly when the students tested speak some English. The tests fail to classify as English proficient students who are fluent in English because they cannot tell the difference between a student who does not know English and a student who does not know the answer or who refuses to answer. Moreover, all test scores, whether they are English proficiency tests or standardized achievement tests, are negatively correlated with SES. There is simply no test made that does not show that relationship.

Indeed, if we simply assume that every so-called LEP student was in fact raised in a lower socioeconomic status English monolingual family, Figure 2 indicates that we should expect about 1/2 of these English monolingual students to never attain English "proficiency." But standardized test scores are not the answer, since Figure 2 also indicates that poor students from English speaking families never achieve parity with all students on standardized achievement tests. Furthermore, there is no way to eliminate inequality in test scores since the tests are periodically normed to produce exactly this outcome. Like a dog chasing its tail, reformers try to eliminate the normal curve, but assess their efforts with tests deliberately constructed to produce a normal curve.

How Long Do Below-Average Students Need Extra Help?
Another disagreement I have with the conclusions of Hakuta et al. is their assumption that children need to be in a special classroom or need special education services if they are below average. Hakuta, Butler, and Witt
apparently believe that students are always helped by special education services, but that is not necessarily the case. It really depends on whether the problem has been accurately diagnosed and what the treatment is. If, for example, an English proficient student is incorrectly classified as LEP simply because the student scores below average on an English proficiency test, the student will undoubtedly be helped if the treatment is after-school instruction or tutoring in English and other subjects that is tailored to their needs. But this is difficult and expensive, and very few school districts in the U.S. do this.

The typical treatment for students who have been diagnosed LEP occurs during the school day so the students receive no additional instruction. The treatments are: (1) a bilingual education program with native tongue instruction if they are believed to be from a Spanish speaking family and there are enough of them to fill a classroom; (2) an ESL pullout program; or (3) a structured immersion program, that is, a self-contained classroom of LEP students taught in English at a slower pace than in the mainstream classroom.

A bilingual education program in Spanish cannot help, and probably harms, a child who does not speak Spanish. Furthermore, such inappropriate treatments do in fact occur as a result of erroneous classifications produced by English proficiency tests. For example, from 1975 to 1996 in New York City, all Hispanic students were forced to take the L.A.B. regardless of their home language and if they scored below the 40th percentile and there were enough to fill a classroom, were placed in Spanish bilingual education classrooms. In fall 1998, I visited a first-grade Spanish bilingual education class in New York City composed only of Hispanic students. During the Spanish reading period, the teacher translated most of what she said in Spanish into English because there were Hispanic students in her class who understood little or no Spanish. They had been assigned to the bilingual program, not because they did not know English, but because they had scored below the 40th percentile on the L.A.B.

In 1996, the NYC school board began to require that newly enrolled Hispanic students be from a home where a language other than English was spoken before they could take the L.A.B. The number of students classified as LEP declined by 20,000 students in New York City when this policy change was implemented. Thus, at a minimum 20,000 Hispanic students were incorrectly classified as LEP solely because they scored below the 40th percentile, and some unknown percentage of them were assigned to a Spanish bilingual education program although they did not speak Spanish. It is hard to imagine how this "special service" could help the English proficiency of these children.

While perhaps not as obvious, an ESL pullout program for a child who is fluent in English can also harm a child. ESL programs take the children out of the mainstream "grade level" classroom for an hour or more a day or a few hours a week, and place them in a small group where they learn basic grammar and concepts that are well below grade level under the assumption that they do not speak English. If the children already know what is being taught in the ESL class, but still need to learn what is being taught in the mainstream classroom during the time they are pulled out for ESL, they will be harmed by the ESL class.

Similarly, a structured immersion classroom is not a beneficial treatment for a child who is fluent in English because like ESL instruction, it is also below grade-level instruction. The teacher teaches content at a slower pace because the students are assumed to not know English. If the students already know English, they will be harmed by this slower pace. In short, special education services can in fact harm students if they do not need the slower pace. This is simple logic that is ignored by Hakuta, Butler, and Witt.

Is a Year in Structured English-Immersion Enough?

What little research there is suggests that although it could take a decade for a student to reach the highest level of English language achievement they are capable of, with students who come to the U.S. at earlier grades reaching it sooner than students who enter in the later grades (Rosell, 2000), all students understand enough English sometime during the first year to be able to comprehend English instruction. I base this conclusion on research conducted in Canada and the U.S. on immersion programs, research conducted in the U.S. and Europe on newcomer centers, my conversations with LEP students in bilingual and ESL classrooms around the U.S., and my conversations with former LEP students in my classes at Boston University.

The studies of French immersion programs in Canada indicate that the English-speaking students, albeit self-selected, eager language learners, understood what the teacher said to them in French sometime during the first semester of the first year. By the end of the second year they were almost equal to French native speakers on many tests (Genesee, 1984; Swain and Lapkin, 1982).

According to Glenn and de Jong (1996), the common European program for immigrant children is to integrate kindergarten children immediately into the mainstream classroom but to provide a "reception" class for one year for those who arrive after the usual age for beginning school. In the reception classes, the focus is on laying the foundation for enrollment in the mainstream classroom. The Europeans have no illusion that the language barrier will be overcome in a year, but they do believe that a year will
provide a solid foundation for older students, and that the language barrier will only be overcome when the immigrant children enrolled in a classroom where they can interact with native speakers of the target language.

These one-year programs are also found in the U.S. under a variety of labels. McDonnell and Hill (1993) found "newcomer" schools for immigrant children in every school district they studied, including the three California school districts, San Francisco, Los Angeles, and Visalia. The length of time for students in the newcomer school was six months to a maximum of one year. McDonnell and Hill describe them as follows:

The newcomer schools in our sample are impressive places; in their clear sense of mission, innovative curricula, professional teaching staff, and links to the larger community, they represent the kinds of schools to which all children, immigrant and native born, should have access... The newcomer schools in our sample all self-contained programs that students attend full-time for one or two semesters (emphasis added), and all but the Los Angeles high school operate in physically separate locations. However, there are a variety of other newcomer models, including ones that students attend for half day and then spend the remainder of the day in mainstream classes. In contrast to the schools in our sample, in which students from across a district are transported to a single site, some districts, such as Long Beach, operate newcomer classrooms on as many as a dozen different campuses. For a description of these other program models see Chang (1990) (McDonnell and Hill, 1993, pp. 97-98).

In addition to newcomer schools, there are one-year immersion programs for kindergarten students all over California and the U.S. In Chelsea, Mass., there are one-year kindergarten immersion programs for Cambodian and Vietnamese students. In New York City there are a number of one-year kindergarten immersion programs (all of them called bilingual) for non-Hispanic LEP students, as well as entire schools for newcomers. One in particular, is the one-year kindergarten immersion program for Chinese students at the Sampson School (P.S. 160) in Brooklyn. In Boston, there is a one-year kindergarten immersion (called bilingual) program for Cape Verdean students at the Mason School. Although Mason School parents have the option of going on to a Cape Verdean "bilingual" program at another school for first grade, very few do that. The conclusion of the teachers and the parents of LEP students at this school is that one year is enough. Within one year, students comprehend enough English to be active participants in the mainstream classroom, although they have a long way to go before they reach their full capacity in English.

I have also had conversations with LEP students in public schools in California, Massachusetts, New York City, and St. Paul, Minn. In most ESL classrooms I have been in, there are one or two students who are working independently because they already know what is being taught. I have taken the opportunity to talk to these students about how long it took them before they could understand what the teacher was saying in English when they entered the school. Those who started in September, having just come from a foreign country, believe they understood what the teacher was saying by the Christmas break. I have also discussed this issue with students in my classes at Boston University who had immigrated to this country as children. None had ever been assigned to a bilingual education class, and all believed they could understand the teacher completely by the end of their first year in an English speaking classroom.

It may be that a few students would be better off staying a little longer than a year in a structured immersion classroom. We simply do not know. What we do know is that we cannot rely on test results such as those presented in the Hakuta, Butler, and Witt report to accurately place or exit students from programs because those standards will result in more than half the students never being classified English proficient, even if that is the only language they know.

This is not just hypothetical, it actually occurs. Table 3 shows the annual reclassification rates for LEP students in California from 1981-82 through 1998-99 (the first year of Proposition 227) using standards such as those in the Hakuta, Butler, Witt report. About 5 percent to 7 percent of LEP students are redesignated English proficient each year in California. If we add up these annual reclassification rates, less than half of a kindergarten cohort that began school in 1992-93 would be redesignated English proficient by the end of their elementary school career in 1998-99, although there is no way the others could not be fluent in English after this time period.

Thus, Proposition 227 is deliberately worded to limit the time period in a separate below-grade level classroom to one year, not because anyone thinks non-English speaking children will have mastered English in one year, but because what evidence there is suggests that sometime during their first year immigrant children will understand enough English so that they will be better off in a grade-level mainstream classroom than in a remedial classroom. Furthermore, if a time limit were not specified in the legislation, more than half of them would never be mainstreamed, no matter how fluent they were in English.
Table 3.
Redesignation Rates for English Learners (Limited-English-Proficient Students) and Cumulative Redesignation Rates for 1992-93 Kindergarten Cohort in California
1981-82 to 1998-99

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of LEP Students</th>
<th># of K-12 Enrollment</th>
<th># of Students Redesignated FEP</th>
<th>% Redesignated FEP</th>
<th>Change from Previous Year</th>
<th>1992 Cohort School Grade</th>
<th>Cumulative % Redesignated FEP w/ Assumption of Same Students in Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998-99</td>
<td>1,422,692</td>
<td>24.7%</td>
<td>106,288</td>
<td>7.6%</td>
<td>0.6%</td>
<td>6th</td>
<td>44%</td>
</tr>
<tr>
<td>1997-98</td>
<td>1,406,166</td>
<td>24.6%</td>
<td>96,545</td>
<td>7.0%</td>
<td>0.3%</td>
<td>5th</td>
<td>37%</td>
</tr>
<tr>
<td>1996-97</td>
<td>1,381,393</td>
<td>24.6%</td>
<td>89,144</td>
<td>6.7%</td>
<td>0.3%</td>
<td>4th</td>
<td>30%</td>
</tr>
<tr>
<td>1995-96</td>
<td>1,323,767</td>
<td>23.8%</td>
<td>81,733</td>
<td>6.5%</td>
<td>0.5%</td>
<td>3rd</td>
<td>23%</td>
</tr>
<tr>
<td>1994-95</td>
<td>1,262,982</td>
<td>23.6%</td>
<td>72,074</td>
<td>5.9%</td>
<td>0.4%</td>
<td>2nd</td>
<td>16%</td>
</tr>
<tr>
<td>1993-94</td>
<td>1,215,218</td>
<td>23.1%</td>
<td>63,379</td>
<td>5.5%</td>
<td>0.4%</td>
<td>1st</td>
<td>11%</td>
</tr>
<tr>
<td>1992-93</td>
<td>1,151,819</td>
<td>22.2%</td>
<td>54,530</td>
<td>5.1%</td>
<td>-0.6%</td>
<td>Kind</td>
<td></td>
</tr>
<tr>
<td>1991-92</td>
<td>1,078,705</td>
<td>21.1%</td>
<td>55,726</td>
<td>5.6%</td>
<td>0.0%</td>
<td>Kind</td>
<td></td>
</tr>
<tr>
<td>1990-91</td>
<td>980,462</td>
<td>19.9%</td>
<td>49,001</td>
<td>5.7%</td>
<td>-1.5%</td>
<td>Kind</td>
<td></td>
</tr>
<tr>
<td>1989-90</td>
<td>861,531</td>
<td>18.1%</td>
<td>53,223</td>
<td>7.2%</td>
<td>-1.2%</td>
<td>Kind</td>
<td></td>
</tr>
<tr>
<td>1988-89</td>
<td>742,559</td>
<td>16.1%</td>
<td>54,482</td>
<td>8.4%</td>
<td>-1.0%</td>
<td>Kind</td>
<td></td>
</tr>
<tr>
<td>1987-88</td>
<td>652,439</td>
<td>14.5%</td>
<td>57,385</td>
<td>9.4%</td>
<td>0.0%</td>
<td>Kind</td>
<td></td>
</tr>
<tr>
<td>1986-87</td>
<td>613,224</td>
<td>14.0%</td>
<td>53,277</td>
<td>9.4%</td>
<td>-1.1%</td>
<td>Kind</td>
<td></td>
</tr>
<tr>
<td>1985-86</td>
<td>567,564</td>
<td>13.3%</td>
<td>55,105</td>
<td>10.5%</td>
<td>0.2%</td>
<td>Kind</td>
<td></td>
</tr>
<tr>
<td>1984-85</td>
<td>524,076</td>
<td>12.6%</td>
<td>50,305</td>
<td>10.3%</td>
<td>-0.1%</td>
<td>Kind</td>
<td></td>
</tr>
<tr>
<td>1983-84</td>
<td>487,835</td>
<td>11.9%</td>
<td>47,503</td>
<td>10.4%</td>
<td>-1.8%</td>
<td>Kind</td>
<td></td>
</tr>
<tr>
<td>1982-83</td>
<td>457,540</td>
<td>11.2%</td>
<td>52,504</td>
<td>12.2%</td>
<td>-3.0%</td>
<td>Kind</td>
<td></td>
</tr>
<tr>
<td>1981-82</td>
<td>431,449</td>
<td>10.7%</td>
<td>57,336</td>
<td>15.2%</td>
<td></td>
<td>Kind</td>
<td></td>
</tr>
</tbody>
</table>


Endnotes
2 While it might seem to be common sense that a child who receives special education services will be better off than one that does not, the most common finding of the research evaluations of special education services such as Title I, Headstart, and bilingual education over the last decade is no effect.
3 The MacMillan test is a standardized achievement test that District B uses as an English proficiency test by establishing its own criterion for "proficiency."
4 The median is that point at which 50 percent of the cases are above and 50 percent are below.
5 This was changed to the 40th percentile in 1989.
6 A home language survey is the first step in identifying a new student as potentially LEP in school districts in the U.S. Typically, a new student will take an English proficiency test only if a language other than English is or was spoken by someone in the home.
7 The highest level of English that a student is capable of is different from attaining parity with native English speakers or a test publisher's standard. Determining the highest level of English an LEP child is capable of requires a sophisticated research design that would attempt to determine an LEP child's intelligence through nonverbal tests and then the standardized test score received by native English-speaking children of that intelligence level. When the LEP child has reached the test score of the native English-speaking students of their intelligence level, they are more or less at the highest level of English they are capable of.

Bibliography
Used for Identifying Children of Limited English Speaking Ability in Texas." Houston, Texas: Education Service Center [Region IV].


---

About the Authors

RALPH E. BEALS, Ph.D., is Clarence Francis professor of economics at Amherst College, where he has served as Acting Dean of the Faculty and Acting President. Among other publications, he is the author of Economic Statistics, an econometric text that has been widely used in public policy programs.

ROGER CLEGG is vice-president and general counsel of the Center for Equal Opportunity in Washington, D.C., where he writes, speaks and conducts research on civil rights issues. A graduate of Rice University and Yale Law School, Mr. Clegg served for 11 years at the U.S. Justice Department, including four years as a deputy in the civil rights division.

WILLIAM A. MEHRENS is professor of educational measurement at Michigan State University. His specializations are the legal issues and the consequences of large-scale assessments. He has published widely in the field, and is a past president of the National Council on Measurement in Education.

SUSAN E. PHILLIPS combines the unique educational background of a doctorate in educational measurement and a law degree. She has published and lectured extensively on legal issues in assessment. As a consultant to many state departments of education, Dr. Phillips has focused on the psychometric and legal issues related to the defensibility of high-stakes assessments.

ROSALIE PEDALINO PORTER is the author of Forked Tongue: The Politics of Bilingual Education (Basic Books, 1990; 2nd edition Transaction Publishers, 1995) and is the editor of READ Perspectives, a refereed journal published by the Institute for Research in English Acquisition and Development (READ) in Washington, D.C. She has lectured on education policy for language minority students on behalf of the U.S. State Department in Bulgaria, Finland, Israel, Italy, Japan and Turkey.

CHRISTINE H. ROSELL is professor of political science at Boston University. She has published four books and more than 50 articles on school desegregation, bilingual education and education policy. Her most recent book, Bilingual Education in Massachusetts: The Emperor Has No Clothes, was published in 1996 by Pioneer Institute of Boston.

ROBERT E. ROSSIER is an independent consultant on English language programs for immigrant students for several California school districts. He has lectured and written on various aspects of bilingual education policy and pedagogy, and served on the National Advisory and Coordinating Council on Bilingual Education in Washington, D.C.