

Published:  
1986

# MIRROR OF LANGUAGE

*The Debate on Bilingualism*

---

KENJI HAKUTA

---

*Basic Books, Inc., Publishers*

NEW YORK

## Chapter 2

---

# *Bilingualism and Intelligence*

"YOU KNOW, I've always wondered. Is it good or bad for children to be bilingual?" This is the question I am most frequently asked. It comes up in my office with students and colleagues; at cocktail parties and during dinner conversations; at professional meetings of scholars and school teachers. My questioners often know of a child whose learning problems have been attributed to bilingualism. They have heard that using two languages in early childhood creates a split personality, a linguistic Jekyll and Hyde. They have also heard that bilingualism is a good thing, that it enriches the intellect, creates a multidimensional view of the world. "Tell me," they demand, "what is the truth?"

It is no wonder that the average person is confused, for scholars have made both claims: bilingualism is bad; bilingualism is good. George Thompson (1952), in a widely used American textbook on child psychology, wrote:

There can be no doubt that the child reared in a bilingual environment is handicapped in his language growth. One can debate the issue as to whether speech facility in two languages is worth the consequent retardation in the common language of the realm. (P. 367)

On the other hand, Canadian researchers Elizabeth Peal and Wallace Lambert (1962) triumphantly drew a contrasting picture of the bilingual as

a youngster whose wider experiences in two cultures have given him advantages which a monolingual does not enjoy. Intellectually his experience with two language systems seems to have left him with a mental flexibility, a superiority in concept formation, a more diversified set of mental abilities. . . . In contrast, the monolingual appears to have a more unitary structure of intelligence which he must use for all types of intellectual tasks. (P. 20)

The primary objective of this chapter will be to look at the literature on "good" and "bad" bilingualism with respect to an elusive psychological construct called "intelligence." Since the turn of the century, psychologists have tried to develop objective instruments for measuring this construct, and researchers have used the performance of bilinguals on such measures as an indication of whether bilingualism is good or bad.

An overview of the hundreds of studies that compare the performance of bilinguals with monolinguals on various measures of intelligence reveals that research in the first half of this century was guided by the question of whether or not bilingualism has a negative effect on intelligence, while more recent work has been concerned with whether or not there is a positive effect. This shift in emphasis is related to the subject populations that were under study. The early work was conducted primarily in the United States with immigrant groups, and the recent work with middle-class populations in Canada and Europe. Although these studies all compare monolinguals with bilinguals, close inspection reveals different motivations behind the studies. The researchers were working under different sociological circumstances. They differed in what moved them to look at the relationship between bilingualism and intelligence in the first place. They chose different methodologies that re-

flected their motivations. And their motivations markedly influenced their interpretations of their findings.

It is difficult to overstate the importance of the *Zeitgeist* in which the scientist works. The importance of understanding this influence is particularly pressing in the case of the social scientist, whose subject matter is so closely connected to his or her own membership in society. Let us begin, therefore, by going back some eighty years to absorb the social context in which the relationship between bilingualism and intelligence began to be investigated in the United States.

#### *The "Old" and "New" American Immigrants*

The first studies of bilingualism and intelligence were not concerned with bilingualism per se. If they considered it at all, they rejected bilingualism (or lack of English ability) as an explanation of intelligence, preferring instead to attribute such differences to racial and ethnic origins.

The motivation for these American studies of the early 1900s was the concern over the changing pattern of immigration from Europe. The Dillingham Commission, set up by Congress in 1907 to investigate the changes, reflected this social trend. It drew a solid distinction between "old" and "new" immigrants, the temporal boundary being set in the early 1880s. The commission lauded the old immigrants from northern Europe, who had dispersed throughout the country and been rapidly assimilated. Its contrasting view of the new immigrants from southern and eastern Europe was characterized by historian Maldwyn Jones (1960) as follows:

This "new" immigration had consisted, [the commission] declared, largely of unskilled male laborers, a large proportion of whom had

come to the United States not as permanent settlers but simply as transients. Almost entirely avoiding agriculture, they had flocked to the industrial centers of the East and Middle West, where they had "congregated together in sections apart from native Americans and the older immigrants to such an extent that assimilation [had] been slow." (P. 178)

What the commission failed to take into consideration, however, was the differences in the length of time the two groups of immigrants had had to settle in their new country. As Jones makes clear, the characterization of the new immigrants is one that applies equally well to the initial wave of both groups. (1960, pp. 177-82).

Coupled with the characterization of the new immigrants as transient and isolated was the view that they were of inferior intelligence. Francis A. Walker (1840-1897), president of M.I.T., wrote,

These immigrants are beaten men from beaten races, representing the worst failures in the struggle for existence. . . . Europe is allowing its slums and its most stagnant reservoirs of degraded peasantry to be drained off upon our soil. (Quoted in Ayres 1909, p. 103)

This characterization of the new immigrants fueled the public outcry for the restriction of immigration of southern and eastern Europeans. The caricature of the new immigrants became an accepted stereotype of these ethnic groups.

The creation of an instrument to measure intelligence went hand in hand with the movement to restrict the flow of the new immigration (Gould 1981; Kamin 1974) though there is debate over the degree to which the testing results were actually employed in the formulation of policy (Samelson 1975; Snyderman and Herrnstein 1983). Following Francis Galton (1890), a number of psychologists in the late nineteenth century were searching for objectively administered measures of intelligence to reflect this most complex of human traits. It would be conve-

nient, they thought, if people could be classified along a single dimension, if "intelligence," like height, were a simple measurement. Then if some measure of this variable called intelligence could be constructed, the measure would be an indicator of a person's worth, and social decisions could be made (and justified) on this basis (Laosa 1984). As Galton, father of the eugenics movement, once wrote,

One of the most important objects of measurement . . . is to obtain a general knowledge of the capabilities of a man by sinking shafts, as it were, at a few critical points. In order to ascertain the best points for the purpose, the sets of measures should be compared with an independent estimate of the man's powers. (1890, p. 380)

The earliest attempts to find measurable capacities linked to intelligence were made in the area of physical characteristics, such as grip strength, lung capacity, and acuity of hearing, which not surprisingly proved unrelated to mental capacity. The critical contribution was made by Alfred Binet, professor of psychology at the Sorbonne, whom the French government appointed in 1904 to devise a method of identifying children who would not benefit from instruction in regular classes and should be segregated for special instruction.

Binet devised a test that included items of some complexity and of varying levels of difficulty. One of Binet's greatest insights was that test items could be arranged with respect to the average age at which children passed them, so that simple observation of a child's performance on these items would permit a general assessment of mental age. The items tapped performance on a variety of skills, including counting coins, repeating sentences, naming the months of the year, noticing pictures with missing parts, and arranging a series of weights.

Binet himself, primarily interested in the assessment and remedial aspects of his work, was quite atheoretical in his approach to intelligence. He was vehemently opposed to the idea

that what his test measured was some fixed entity, unmodifiable through experience. As Leon Kamin (1974) remarks, "It is perhaps as well that Binet died in 1911, before witnessing the uses to which his test was speedily put in the United States" (p. 5).

In 1910, H. H. Goddard, who was director of the Vineland School for Feeble-Minded Girls and Boys in New Jersey, translated the Binet test into English for use in the United States and made it available for use in assessing the intelligence of immigrants. In one study, Goddard (1917) took the English-language version of the Binet test to Ellis Island, the point of entry for newly arrived immigrants. In testing thirty adult Jews through an interpreter, he assessed twenty-five of them as "feeble-minded." Regarding their performance on a word-fluency section of the test, Goddard wrote:

What shall we say of the fact that only 45 per cent can give sixty words in three minutes, when normal children of eleven years sometimes give 200 words in that time! It is hard to find an explanation except lack of intelligence or lack of vocabulary and such a lack of vocabulary in an adult would probably mean lack of intelligence. How could a person live even fifteen years in any environment without learning hundreds of names of which he could certainly think of 60 in three minutes? (P. 251)

The fact that his test found over three-quarters of this group feeble-minded did not raise doubts in Goddard's mind about the validity of the test, even though he had administered it under circumstances that were unfamiliar to and most likely traumatic for the new arrivals. Rather, Goddard took these assessments to be true measures of his subjects' intelligence and concluded that "we are getting now the poorest of each race. This makes them a highly selected group at the start" (p. 266). Goddard's recommendation, based on this research, was that "if the American public wishes feeble-minded aliens excluded, it must demand that Congress provide the necessary facilities at the ports of entry" (p. 271).

Following Goddard's lead, there was an almost immediate explosion of new tests and research. (By the 1930s, a bibliographic listing of research studies on testing in America was 251 pages long, and a "bibliography of bibliographies" itself took a full six pages [Goodenough 1946]). Lewis Terman, a professor of psychology at Stanford University, was perhaps the strongest advocate of the tests. He extended the Binet test to include older children and adults and refined the method for determining the intelligence quotient (IQ). His version of the test, for example, included the now-familiar multiple choice format, such as "Napoleon was defeated at: Leipzig / Paris / Verdun / Waterloo." Also included were sentences containing absurdities to be noticed and explained, such as "Yesterday the police found the body of a girl cut into 18 pieces. They believe that she killed herself" (Terman 1916, 1926). His revisions of the Binet test (the Stanford-Binet) came to be the prototype IQ test, an industry standard against which all new tests had to be compared.

The outbreak of World War I made possible testing on a large scale. Professor Robert Yerkes, of Harvard University, in collaboration with Terman and Goddard, persuaded the United States Army to test some two million draftees, purportedly to aid in classifying the new recruits. They constructed two group tests, one intended for those who could read and write English (Alpha) and one for illiterates and "foreigners" (Beta), who were given instructions in pantomime. Since the soldiers represented a variety of foreign nationalities, it became possible to make group comparisons by racial origin.

Famous among the popularizers of these data was Carl C. Brigham, who analyzed them in a book titled *A Study of American Intelligence* (1923). Of prime interest for Brigham was the pattern in test performance among the new immigrant groups. While the "foreign born white drafts" who had been in the United States for over twenty years were comparable to "native born white drafts," those with shorter years of residence in the

United States fared much more poorly, a fact that Brigham interpreted in the following way:

Migrations of the Alpine and Mediterranean races have increased to such an extent in the last thirty or forty years that this blood now constitutes 70 percent or 75 percent of the total immigration. The representatives of the Alpine and Mediterranean races in our immigration are intellectually inferior to the representatives of the Nordic race which formerly made up about 50 percent of our immigration. (P. 197)

Statistical problems in this analysis aside (see Gould 1981), a major alternative explanation stood in the way of this conclusion. The number of years of residence in the United States is obviously related to the knowledge of English and the level of acculturation to American society. It is thus directly related to the ability to answer correctly such questions as "Why should a married man have his life insured?" (Alpha Test 3, Item 13).

Brigham's response to this problem can be seen as the origin of the so-called "language handicap of bilinguals" issue. Brigham was an uncompromising hereditarian, who believed in the unmodifiability of native intelligence. Intelligence tests measured native intelligence, and nothing, not even unfamiliarity with the language, attenuated their results. The issue of language handicap, then, as it was originally raised, had to do with a measurement issue, of whether persons who happened to be bilingual were hindered by their lack of control of the language of the test.

In arguing that bilinguals did not suffer from a language handicap in taking intelligence tests, Brigham separated the new immigrant groups into those who had taken the Alpha (for literates) and those who had taken the Beta (for illiterates and foreigners). He showed that the pattern of decreasing scores with recency of immigration held not just for those who took the Alpha test, which might be expected if there were a language handicap, but also for those who took the Beta test,

which presumably did not depend on knowledge of English (p. 102).

Brigham had little sympathy for the possibility that attitudes toward testing and other cultural factors might have significantly influenced the results.

It is sometimes stated that the examining methods stressed too much the hurry-up attitude frequently called typically American. The adjustment to test conditions is a part of the intelligence test. We have, of course, no other measure of adjustment aside from the total score on the examinations given. If the tests used included some mysterious type of situation that was "typically American," we are indeed fortunate, for this is America, and the purpose of our inquiry is that of obtaining a measure of the character of our immigration. (P. 96)

Apparently, Brigham assumed that test-taking ability is part of native intelligence.

Perhaps the most telling evidence of Brigham's hereditarian attitude is his attempt to rule out the language handicap by dividing the Nordic immigrants into those from English-speaking countries and those from non-English-speaking countries. When these groups are compared, a clear difference emerges in favor of the English-speaking Nordics. The analysis showing the language handicap is as clear-cut as any of those in Brigham's book. But rather than dwell on this obvious contribution of experience to the test scores, Brigham chose to dismiss it by saying, "There are, of course, cogent historical and sociological reasons accounting for the inferiority of the non-English-speaking Nordic group" (p. 171). He then compared the non-English-speaking Nordic group with the Mediterranean group and found a difference in favor of the Nordics, "a fact which clearly indicates that the underlying cause of the nativity differences we have shown is race, and not language" (p. 174).

*Bilingualism and the Nature-Nurture Issue*

In contrast to the hereditarian view of bilingualism that emphasized the genetic quality of groups who happened to be bilingual, psychologists of the experiential orientation stressed the role of the environment of the bilinguals. This tension over "nature versus nurture," a controversy that can be found running through much of academic psychology, is to a large extent a matter of emphasis. Very few hereditarians deny any contribution of the environment, and few experiential psychologists deny the relevance of a person's genetic endowment. Rather, the difference lies in their beliefs about the extent to which traits such as "intelligence" can be modified through experience.

During the early part of the twentieth century, the struggle between the two positions was symbolized by the ongoing debate (National Society for the Study of Education 1928, 1940) between psychologists at the Iowa Child Welfare Station at the University of Iowa (George Stoddard and Beth Wellman) and those at the University of Minnesota (Florence Goodenough) and at Stanford University (Lewis Terman). The Iowa emphasis on experience is reflected in a textbook by Stoddard and Wellman (1934), in which they acknowledge that "the great bulk of mental ability as measured by tests comes as a direct inheritance" but emphasize that "the real question concerns the amount of variability which can still be effected by later influence" (p. 170).

The Minnesota/Stanford attitude is best characterized as unforgiving, preferring an explanation based on heredity even when an alternative account based on the environment is possible. An illustration of this attitude can be found in the following argument, provided by Goodenough (1940) to explain the low intelligence of people of an inbred, "backward mountain community" called Colvin Hollow:

Given two centuries of social anemia, during which time all the ablest members of the group have been continuously drained away, leaving only the intellectual and volitional weaklings to interbreed and reproduce their kind, *need we seek further for an explanation of the state of educational backwardness and intellectual degeneracy found?* (P. 329, emphasis added)

The question of bilingualism and intelligence must be seen in the context of these conflicting approaches. For hereditarians, bilingualism was irrelevant to the major focus of study. Eager to show that intelligence was based on heredity, they were not the ones to argue that poor performance on intelligence tests could reflect a language handicap. Rather, it was those researchers with the experiential orientation who considered bilingualism—a learned characteristic—to be the cause of low intelligence.

#### *The Hereditarians and the Language Handicap*

Arguing for the genetic inferiority of bilinguals therefore required the hereditarians to demonstrate that the bilinguals did not suffer from a language handicap when their intelligence was being tested. Lewis Terman's students played a central role in this debate.

Terman himself began framing the debate in 1918, when he reported that for both monolingual English-speaking children and children of Portuguese and Italian immigrant families, a simple vocabulary test was a good reflection of mental age as measured in an IQ test. Terman reported that after children had been in school three or four years, their vocabulary and mental age scores correlated as well for the foreign children as it did for the Anglo children. He failed to note that the high correlation for both groups might arise from the fact that both measures reflect the degree of knowledge of English.

In 1922, Kimball Young published an influential article in *Scientific Monthly*, summarizing a set of arguments against the language handicap. In one argument, he held that the inferiority of the foreign children persisted even after the children had had a chance to learn English. In support, he cites a Master's Thesis directed by Terman, in which southern European children were followed up over a two-year period; they remained behind American children of northern European stock.

Another form of argument, supported by Young's own dissertation, was that verbal tests (the Army Alpha) are a better predictor of school performance (as judged by children's grade level relative to their age, by teachers' estimates, and by school grades) than nonverbal tests (the Army Beta). From this, Young drew the conclusion that "the asserted language handicap under which the foreign children are supposed to labor does not exist, at least so extensively as imagined" (p. 428). Young apparently was reluctant to consider the possibility that school performance is dependent on English skills, which are better measured by the Alpha. As a contemporary critic of Young wrote, "A teacher's estimate of a child's intelligence will unquestionably be influenced by the child's ability to use the English language, and, of course, all the child's school work is conditioned by his ability to understand and make use of English" (Pintner 1923, p. 292).

Such dissenting opinions, however, represented a minority view. The majority opinion is reflected in the review of Young's dissertation that appeared in *The Journal of Educational Psychology*:

The study sheds a bright light on the question of the part played by the language difficulty in the differences among racial groups, repeatedly found in the intelligence scores. By correlations between the several sorts of data, Mr. Young shows very conclusively that the language factor is by no means as great as is commonly believed, and that the differences in scores [between racial groups] is much more largely one of native intelligence. This constitutes a genuine contribution. (Kelly 1923, p. 256)

Young's article also cites data from Japanese and Chinese immigrant children, who generally tested better than Italians and Portuguese and almost on a par with Americans. Assuming that European languages are more similar to English than are the Oriental languages, he argued that "surely the language handicap is of greater potency in the Oriental than in the European" (p. 430). In retrospect, this was probably one of the better arguments advanced by Young against the language handicap; in fact, it is enjoying a revival among contemporary researchers who argue that it is cultural, not linguistic, differences that matter (Troike 1981).

Young's arguments notwithstanding, the inevitable evidence for an English language handicap soon began to surface. Pintner (1923), for example, constructed a "Non-Language Test," which he administered along with the National Intelligence Test, a group test derived from the Army Alpha, to foreign-born children. He found that the children fell considerably behind national norms on the NIT but at national norms on the Pintner Non-Language Test. Margaret Mead (1927) gave the Otis Group Intelligence Scale to sixth- to tenth-grade Italian immigrant children. She found higher IQ scores both as a function of the amount of English spoken at home and as a function of the length of residence in the United States.

Evidence for the language handicap was soon emerging even in Terman's own backyard. Darsie (1926), his own student, administered the Stanford-Binet to 570 Japanese-American children in California. His results were quite straightforward: "Japanese children as compared with American show a mean retardation of 14.25 months in reading, 12.5 months in language, 1.75 months in arithmetic, and 6.0 months in general information. In spelling they average 2.75 months above American children" (p. 86). The more the skills tapped involved English, the larger the discrepancy between English-speaking and Japanese children (the one exception being spelling, which Darsie dismissed as due to the "acute visual perception and sustained attention" of the Japanese [p. 33]).

On the whole, Darsie was forced to admit that the foregoing analysis . . . conclusively establishes the essentially linguistic character of the Binet scale" (p. 59). In his conclusions, however, are to be found the germs of the hereditarian response to the problem posed by the language handicap. The argument goes full circle: "It must not be overlooked, however, that the existence of a pronounced language handicap may itself be indicative of lack of capacity to master the language adequately" (p. 84). Since children of northern European stock apparently have less difficulty mastering English, Darsie concluded (while admitting to the closer linguistic affinity), they must be of superior intelligence.

This line of argument was perfected by Florence Goodenough (1926), who summarized data on the persistence of the foreign language in the homes of immigrants of different nationalities. She showed a negative relationship between the amount of foreign language used in the home and the median IQ of the groups. The less foreign language they used (and the more English), the higher their IQ. Simple correlations never establish causality (a basic principle of statistical inference, which Goodenough surely knew and probably taught), but Goodenough was willing to rest her case:

This might be considered evidence that the use of a foreign language in the home is one of the chief factors in producing mental retardation as measured by intelligence tests. A more probable explanation is that those nationality groups whose average intellectual ability is inferior do not readily learn the new language. (P. 393)

Thus was created the party line of the hereditarians: the language handicap of foreign children in intelligence testing is minimal, so what the tests indicate is that these children are from inferior genetic stock. Even if the language handicap does impede performance, that does not belie the validity of the tests, because the language handicap is itself a result, rather than a cause, of inferior intelligence. Such were the dark

beginnings of the term *language handicap* in the study of the bilingual.

*The Experiential View of Bilingualism*

New technologies in an industrial society are shrouded by an aura that often makes them resistant to critical evaluation. There is no question that intelligence tests in the early 1900s were such an enshrouded technology. American psychologists generally considered intelligence tests to be their ticket of admission to the brotherhood of the natural sciences. If one considers the "hard" sciences to be defined by rigorous methodology, careful measurement, and quantification (rather than by the questions one asks), psychometrics certainly provides room for such activities. Because of the respect American psychologists had for intelligence testing, the debate centered almost exclusively on whether differences among individuals and groups on these measures reflected heredity or experience, and not whether the measures themselves were adequate and equivalent for all the individuals tested.

In this context, if you tested bilinguals on a measure of verbal intelligence, and if the bilinguals showed inferior performance, you were bound to one of two conclusions. You could conclude, as the hereditarians did, that the bilinguals were genetically inferior. Or you could conclude that bilingualism caused some kind of mental confusion, resulting in the poor development of verbal skills. The possibility that the tests were limited in their ability to measure intelligence in this population was not considered.

One of the more perversely humorous examples of this underlying faith in the tests comes from the conclusion that A. J. Mitchell (1937) drew from a very well intentioned study com-

paring Mexican-American children's performance on an English intelligence test and on a Spanish translation of it. He found that there was consistently better performance on the Spanish version, which he regarded as a truer estimate of the children's intelligence. Rather than conclude that the English test was useless, however, Mitchell recommended that "thousands of cases" be tested in both languages for each grade and that a "correction figure" be calculated to adjust scores obtained from testing in English. No matter what their motivations were, the psychological researchers in those days were committed to the idea that these tests really measured intelligence.

Within the psychometric tradition, the earliest work in support of the negative effects of the experience of bilingualism, widely cited in the American literature, can be found in the British journals, especially in relation to Welsh-English bilinguals in Wales. Frank Smith in 1923 reported in the *British Journal of Psychology* a study comparing monolingual and bilingual third- and seventh-graders in the same school. He found that the monolinguals were better in tasks involving dictation, sentence-forming, and composition in English. He also reported a two-year longitudinal follow-up on similar measures, in which he found more improvement over time for the monolinguals than for the bilinguals. Smith concluded that "bilingualism may yet be shown to be no intellectual disadvantage in the young; but the tests described in this paper clearly support the view that under present methods it is a positive disadvantage" (p. 281).

The following year, Saer (1924) reported a more systematic study of Welsh-English bilingual and monolingual children aged seven to fourteen, in which the measures included the Stanford-Binet, a test of "dextrality," and vocabulary and composition tests. Saer divided the subjects into rural and urban samples and found that there were differences between bilinguals and monolinguals from the rural areas but not between those from the urban areas. In a second study, Saer found a

similar pattern of differences in a comparison of bilingual and monolingual university students from rural and urban areas.

Of prime interest for our purposes is not the result but the interpretation of the apparent differential effects of bilingualism on children in rural and urban environments. Saer apparently was oriented toward emotional and "psychodynamic" explanations, and he claimed that for the urban bilingual children, "any emotional conflict between the use of Welsh and English that may arise is resolved by the child at an early age" (p. 37). On the other hand, for the rural child, "since the Welsh [language has] for him a high affective tone, and since the cathartic influence of play does not operate, for he uses Welsh in play, a conflict must arise between his self-regarding sentiment or positive self-feeling and his negative self-feeling or his instinct for submission" (p. 37).

Among American researchers, Yoshioka (1929) advanced the interpretation that the experience of bilingualism had negative consequences on intellectual development. He conducted a small study of Japanese-American children, to whom he administered English and Japanese versions of the National Intelligence Test (for which norms were available in English and Japanese). His conclusion was that "bilingualism in young children is a hardship and devoid of apparent advantage, because bilingualism appears to require a certain degree of mental maturation for its successful mastery" (p. 479).

Yoshioka's research was followed up by Madorah Smith (1931, 1939), the most influential proponent of the negative consequences of bilingualism, whose studies were extensively cited in later literature (see, for example, McCarthy 1946; Thompson 1952).

Smith received her doctorate at the Iowa Child Welfare Research Station at the University of Iowa, the center of research oriented toward experiential influences on intelligence. In her dissertation, published in 1926, Smith had pioneered a method of analyzing free speech utterances of young monolingual chil-

dren. After obtaining her degree, Smith moved to the University of Hawaii and began applying her method to the speech of bilingual children from the wide variety of language backgrounds represented on the islands (Smith 1939). She studied the speech of children between the ages of two and six from Japanese, Chinese, Korean, Filipino, Hawaiian, and Portuguese backgrounds and compared them with the Caucasian norms that she had developed for her dissertation.

We will return to Smith's extensive study in the next chapter, when we contrast this line of work with that of the linguists who focused on the intricacies of language rather than on differences between individuals. For now, suffice it to say that when she compared her bilingual samples with her monolingual sample from Iowa on a variety of measures of language, the bilinguals showed inferior performance. Smith concluded that "an important factor in the retardation in speech found in the preschool population is the attempt to make use of two languages" (p. 253). This conclusion, implicating the bilingual experience rather than the genetic quality of the children who happened to be bilingual, is quite different from the kinds of conclusions the hereditarians would have drawn from the same data.

Smith continued her crusade against early childhood bilingualism in a study (1949) of preschool children of Chinese ancestry in Hawaii, who apparently were English-dominant but spoke some Chinese at home. She translated into Chinese a vocabulary test she had developed in Iowa and administered both versions to these children. She found that the vocabulary scores of these bilingual children in both languages were below the monolingual norms, although when the scores from the two languages were added together, they were comparable. She concluded that "it would seem unwise to start any but children of superior linguistic ability at a second language unnecessarily during the preschool years" (p. 309).

Smith's line of argument was followed up by Anne Anastasi,

professor of psychology at Fordham University, much of whose career has been devoted to arguing the "fallacies of 'culture-free' testing and of attempts to assess innate potential" (Anastasi 1980, p. 27). In one study (Anastasi and Cordova 1953), Puerto Rican children (ages eleven to fifteen) in New York City were given Cattell's Culture Free Test in English and Spanish versions. The test was nonverbal, "all items being perceptual or spatial" (p. 5). Anastasi and Cordova found that the language of test administration made no difference. Their subjects performed below the norms in both languages. Anastasi attributed the poor performance to the fact that these children's bilingualism "appears to be of the bifurcated variety, the children's mastery of either language being restricted and inadequate" (p. 13). It is entirely possible to argue that the data had no bearing on bilingualism, since the test itself was nonverbal in nature. Nevertheless, while acknowledging the importance of other factors, Anastasi maintained that bilingualism was the major villain:

Among the reasons for [the poor test performance] are the very low socio-economic level of the Puerto Rican children, their bilingualism which makes them deficient in both languages, their extreme lack of test sophistication, and their poor emotional adjustment to the school situation. In so far as this maladjustment itself appears to have arisen from the children's severe language handicap during their initial school experiences, a solution of the language problem would seem to be a necessary first step for the effective education of migrant Puerto Rican children. (P. 17)

The early history of research into bilingualism and intelligence in the United States is thus convoluted. The initial research concerned the new immigrants, who performed poorly on tests of intelligence. The hereditarians argued that this poor performance reflected inferior genetic stock, not a language handicap in test-taking. As the evidence mounted that bilinguals were operating under a handicap, the hereditarians interpreted this handicap itself to be the result of innately inferior

intelligence. On the other hand, the experientially oriented psychologists took the language handicap in bilinguals to be the result of experience, specifically the experience of being exposed to two languages. In either case, the concept of language handicap, originally construed as a variable related to test-taking factors, came to be a trait of the bilingual individual's mind, whether based on experience or on genetic quality.

#### *A Positive View of Bilingualism*

In Canada, a different set of sociological events surrounded the study of bilingualism, particularly in the Montreal area. In the 1960s, it was becoming increasingly clear that with the rise in the political status of the French language (confirmed by the *Official Languages Act of 1968-69*, which granted equal status to English and French at the level of the federal government), bilingualism was the wave of the future and essential to political power. Parents, especially middle-class parents, were beginning to be concerned with making their children into bilingual citizens. Yet there was also concern, in large part due to the earlier American research, that bilingualism could have harmful effects on children's intellectual development.

It was in this context that Elizabeth Peal and Wallace Lambert conducted their influential study (1962). In their monograph, they reviewed the earlier studies of the language handicap of bilinguals and discredited their validity. These studies, they argued, failed to take into account the fact that the bilingual and monolingual subjects came from different socioeconomic backgrounds. In most cases, bilinguals from poor backgrounds were compared with monolinguals from higher social classes. Moreover, many of the earlier studies did not adequately ensure that their subjects were truly bilingual.

Peal and Lambert's discussion of the early literature was kept

at a purely methodological level, steering clear of the issues of hereditarian versus experiential biases that loomed over that research. Their methodological concerns led to a criterion for sample selection that became standard for subsequent research. They drew a distinction between true, "balanced bilinguals," who are proficient in both their first (L1) and second (L2) languages, and "pseudo-bilinguals," who for various reasons have not attained age-appropriate abilities in their second language. As Peal and Lambert wrote, "The pseudo-bilingual knows one language much better than the other, and does not use his second language in communication. The true (or balanced) bilingual masters both at an early age and has facility with both as means of communication" (p. 6). The bilinguals in their sample were all judged to be equally good in their two languages, on the basis of relatively equal performance on language tasks (including a vocabulary test) in both languages, as well as on subjective self-ratings of their ability in the two languages.

Their bilingual and their monolingual subjects were all ten-year-old children from the same French school system in Montreal. The two groups were equivalent in their measures of socioeconomic status, both solidly middle class. They were administered both verbal and nonverbal measures of general intelligence. Contrary to the findings of previous research, the bilingual children performed reliably better than the monolinguals on both the verbal and the nonverbal measures. The bilingual children's superiority in nonverbal tests was more clearly evident in those subtests that required mental manipulation and reorganization of visual patterns, rather than simple perceptual abilities. A statistical analysis of the structure of the relationship between the different measures indicated that the bilinguals were superior to the monolinguals in concept formation and in tasks that required a certain mental or symbolic flexibility. Thus originated the claim that bilinguals enjoy a certain advantage in "cognitive flexibility" over their monolingual counterparts.

Ever since Peal and Lambert's study, researchers in bilingualism and intelligence have been careful to select subjects who fit some criterion of balance between their two languages. A wide variety of tasks have been given to groups of bilingual and monolingual children of various ages. Peal and Lambert's results have generally been replicated with children in western Ontario and other parts of Canada (Liedtke and Nelson 1968; Bain 1974; Cummins and Gulutsan 1974), Switzerland (Balkan 1970), Israel (Ben-Zeev 1977a), South Africa (Ianco-Worrall 1972), and even the United States (Ben-Zeev 1977b; Duncan and De Avila 1979). These studies are based primarily on middle-class populations. Overwhelmingly, they claim that bilingualism has positive effects (see a recent review by Rafael Diaz [1983]).

These recent studies suggest the following conclusion: take any group of bilinguals who are approximately equivalent in their L1 and L2 abilities and match them with a monolingual group for age, socioeconomic level, and whatever other variables you think might confound your results. Now, choose a measure of cognitive flexibility and administer it to both groups. The bilinguals will do better.

#### *Some Methodological Problems*

What is wrong with the above conclusion? To a rigorous experimental psychologist, such a study has several weaknesses in its method. In order to see what they are, let us indulge in an intellectual exercise and conjure up the ideal experimental design to study the relationship between bilingualism and intelligence.

You begin by taking a random sample of individuals and assigning them randomly to either an experimental group or a control group, thereby controlling for any background "noise"

in sampling. You test both groups before their treatment, to ensure that they do not differ on your measures of cognitive flexibility. The experimental group is then placed in an environment that fosters bilingualism, while the control group remains in a monolingual environment. Once the treatment has had time to take effect—that is, once the subjects in the experimental group have become balanced bilinguals—you administer your dependent measures. As a good experimentalist, you make sure that the person who administers the dependent measure does not know whether the subject being tested is in the treatment or the control group, because no matter how well intentioned the experimenter may be, he or she can bias the outcome of the study if this procedure is not followed. And, lo and behold, you find a difference in favor of bilinguals. Under these ideal conditions, you could reasonably conclude that bilingualism causes cognitive flexibility. You could also go on to speculate about why you got this result and set up various other experimental conditions to test your hypotheses.

In what ways do the current studies of bilingualism and intelligence deviate from this ideal? Let me indicate their shortcomings by describing what I consider to be one of the best studies in this area, conducted by Bruce Bain and Agnes Yu (1980).

The researchers attempted to test the effect of raising children in a bilingual home. Specifically, they were interested in the ability of bilingual children to use language to help direct their thoughts and actions. Bain and Yu placed advertisements in school and university community newsletters in Alsace (France), Alberta (Canada), and Hong Kong asking for parents of newborns to volunteer in a study of “unilingual speech acquisition” or of “bilingual speech acquisition.” From those who responded to the advertisements, they chose thirty sets of parents who would raise their newborns bilingually and sixty sets of parents who would raise their newborns monolingually. Each bilingual family was paired with two monolingual coun-

terparts, one for each of the two languages. In each family, at least one parent had a university degree “and was either a practising school teacher, graduate student, or university professor” (p. 306). When the study began, all the infants were between six and eight months old.

Bain and Yu visited the homes of the subjects at the beginning of the study and offered instruction on how to raise the infants in the monolingual environments. As Bain and Yu describe their procedure,

It was emphasized that regardless of who initiates an exchange, be it the child or a parent, and regardless of which language or mixture of languages the child initially uses, each parent was to conduct all exchanges with the child in his or her designated language. We demonstrated how the parents should set up specific language zones in their home. (P. 307)

The parents of the monolingual children were given a general tutorial on language and cognitive development.

When the children were twenty-two to twenty-four months old, and again when they were between forty-six and forty-eight months old, they were given a variety of cognitive tests devised by the Russian psychologist Luria (who followed the tradition of Vygotsky, described in chapter 4). For example, in one task the experimenter hides a marble under one of several cups, and the child is told which cup it is under. In another task, the child is to follow such instructions as “When the red eye goes on, say ‘squeeze,’ and squeeze the ball. Say it and do it.” The results showed that at twenty-two to twenty-four months, there were no differences between the groups, but at forty-six to forty-eight months, the bilinguals reliably outperformed the monolinguals. They suggested to Bain and Yu the positive effects of rearing children bilingually.

Bain and Yu’s study is a noble attempt to show the effect of bilingualism on cognitive development. They studied the subjects longitudinally, and it is of interest that the two groups did

not differ at twenty-two to twenty-four months, suggesting that there were no initial differences between the groups and that the bilingual treatment brought about the difference at forty-six to forty-eight months. They also drew the bilinguals and monolinguals from a similar occupational class. Thus far, they approximate quite well the experimental design of my fantasy.

The fantasy falls apart, however, when we consider that the children were not randomly assigned to the two groups. Of critical importance is the question of who decided whether the subjects would be in the bilingual group or the monolingual group. If Bain and Yu had had the power to do the assigning, then we could rest reasonably assured that the experiences of the bilingual children caused them to perform better on the cognitive tests, because they would have been selected on a random basis. But that was not the case. In the context of this study, the parents who responded to the bilingual advertisement had presumably already decided to raise their children in this manner. The bilingual subjects, then, were raised by a set of parents who were probably interested in the language heritage of their children and were perhaps considerably different from parents who (whether they had the option or not) raised their children as monolinguals.

Having dealt a painful bruise to Bain and Yu's study, let me turn to the rest of the studies, which suffer from additional flaws. The control imposed by Bain and Yu, of testing children both before and after they have had the chance to become bilingual, is an important feature that has been neglected in other studies in this area. Because of the cross-sectional nature of the studies, we are unable to infer the direction of cause and effect. Peal and Lambert themselves acknowledged that "one may ask whether the more intelligent children, as measured by nonverbal intelligence tests, are the ones who become bilingual, or whether bilingualism itself has a favorable effect on nonverbal intelligence" (1962, p. 13). As the teacher in elementary

statistics repeats over and over, correlation does not imply causation.

An even more serious problem was raised very early on by John Macnamara (1966). Regarding Peal and Lambert's method of selecting only balanced bilinguals, he wrote, "It is extremely likely that in selecting for the bilingual group native French-speakers who had become balanced bilinguals, the authors selected children who on the whole were highly gifted and had a flair for language learning. So any linguistic comparison between these children and the monoglot was probably biased in favour of the former" (p. 21).

Macnamara's objections, and even more vehement reiterations of these arguments by MacNab (1979), have gone virtually unrecognized, as if the problems do not exist (but see Lambert and Anisfeld [1969] and Cummins [1976] for discussion of this issue). The paradigm had shifted in the science of bilingualism and intelligence, and the *Zeitgeist* allowed for studies showing positive relationships.

One recent attempt at addressing the methodological problems pointed out above can be found in a study that Rafael Diaz and I conducted in New Haven (Hakuta and Diaz 1984; Hakuta 1984b). We decided to try to circumvent the problem of between-group comparison by looking at the effects of bilingualism *within* a group of bilinguals. We reasoned that if bilingualism has a positive effect on measures of cognitive ability, then we should be able to show that those who are more bilingual do better on these measures than those who are less bilingual. Furthermore, we decided to employ a longitudinal design, so that we could look at the effects over time.

Our sample included more than three hundred Puerto Rican elementary school children in the bilingual education program in the New Haven public schools. All the subjects were from extremely poor home backgrounds, and all were considerably more proficient in their Spanish than in their English (the criterion for inclusion in the bilingual program). Their Spanish can

be interpreted as their verbal ability in their native language, while their English reflected their degree of bilingualism.

The most important measures of cognitive ability were (1) Raven's Progressive Matrices, a nonverbal test of intelligence used by Peal and Lambert and found to be related to bilingualism, and (2) a measure of "metalinguistic awareness," or the ability to reflect on and evaluate the forms of language (which was administered in Spanish). The critical question was how their abilities in Spanish and English (as measured by a vocabulary test) would bear on these measures.

In brief, it turns out that Raven's is positively related to the degree of bilingualism (ability in English), while the measure of metalinguistic awareness is more strongly related to the native language ability (ability in Spanish). In one sense, it appears that we have support for the notion that bilingualism and nonverbal intelligence are related in a positive way, as found in the more recent studies. It is bothersome, however, that metalinguistic awareness is only weakly related to bilingualism. The difficulty with this pattern of results is that it is not theoretically neat. The most logical route for bilingualism to have an effect on intelligence is through language. One can easily imagine a mechanism in which bilingualism first results in verbal flexibility (reflected in metalinguistic skills), which then generalizes to nonverbal skills. Our results, however, suggest that bilingualism might have an effect on nonverbal intelligence but less of an effect on metalinguistic awareness.

Our current interpretation is that bilingualism in this group bears little relationship to performance on these measures. Rather, we think that English and Raven's are related most strongly because English is the skill that children are concentrating on learning in school, and Raven's is a good measure of overall ability of children to learn in school. There is the distinct possibility, however, that the effects of bilingualism have not had the chance to appear in this population because they are nowhere near being balanced bilinguals. If a certain threshold

level of competence in both languages is required for positive effects to show (Cummins 1976), it may be that this population has not yet crossed the threshold.

We do have some indications of what would happen if our subjects were followed up for several more years and if their bilingualism were allowed to develop (this is in practice impossible, because the policy of transitional bilingual education in the United States means that children are mainstreamed into monolingual English classes as soon as their English is sufficient for survival, and their Spanish then becomes unsupported). Our longitudinal analysis shows that with increasing years of exposure to English, the relationship between English and Spanish becomes substantial (by the end of three years, the correlation between abilities in the two languages approaches .70). This means that, as we follow these students, assuming that English remains related to Raven's and Spanish to metalinguistic skills, these two skills will converge. If subjects are sampled when that occurs, it will appear that the ones who are more bilingual will do better on both metalinguistic and nonverbal measures of cognitive ability. What the study reveals, in essence, is that the way in which sampling is conducted and the way in which bilingualism is defined are going to determine much of the pattern of results.

#### *Reconciling Differences*

Having shown the limitations of the methodologies of current studies, we must still try to account for what appear to be contradictory conclusions about the effects of bilingualism. The effects, negative or positive, correlate with the methodology used. When bilinguals are unselected and come from lower socioeconomic backgrounds, negative effects are found. When

bilinguals are selected for balanced bilingualism and come from middle-class backgrounds, positive effects are found. It is a serious puzzle for anyone seeking consistency in the world. What should be the locus of the resolution?

A scientist is most comfortable staying at the level of the methodology of studies. Peal and Lambert offered a methodological refinement in 1962, in terms of selection of monolingual and bilingual subjects. Through rigorous selection of balanced bilinguals from similar socioeconomic groups, they found positive effects. The implication was that if only the earlier studies had controlled for such factors, they would have found the same positive results.

Jim Cummins (1976) argues along the same lines when he claims that subtractive bilingualism, in which the second language replaces the first language, results in negative effects, while additive bilingualism results in positive effects. He speculates that there must be a "threshold" effect, requiring a certain level of competence in both languages before the positive effects can appear. The earlier studies showing negative effects, by virtue of not selecting for balanced bilinguals, included cases of subtractive bilingualism. My own study just described above, of course, also falls in this camp of attempts at methodological refinement.

Such methodological explanations tend to abstract the bilinguals away from the social conditions in which they live and to focus only on their degree of bilingualism. It so happened that the immigrants with low socioeconomic status were not so bilingual, while the post-Peal and Lambert bilinguals of middle-class background had high degrees of bilingualism.

We must, however, attempt to look beyond our methodological quibbles. It is easy to forget that methodologies are not magical entities that exist independently of researchers. *Researchers create and choose their methodologies.* I suggest that the choice of methodology—of whether to use unselected bilinguals or balanced bilinguals, for example—reflects both the

world view of the researcher and the social status of the researcher relative to the subject population. The researcher sits at the control panel, as it were, and manages the knob of methodology, which in turn determines the pattern of results obtained. This image is a bit too strong, even for my most cynical moments; researchers obviously do not so consciously manipulate their results. Nevertheless, in the long run, a full account of the relationship between bilingualism and intelligence, of why negative effects suddenly turned into positive effects, will have to examine the motivations of the researcher as well as more traditional considerations at the level of methodology and the mental composition of the bilingual individual.

Joshua Fishman (1977) diagnosed the problem accurately when he wrote:

My own socio-historical perspective (bias?) leads me to doubt that answers . . . can be found by better controlled experiments, which in essence, cannot explain shifts in social climate that take place across a decade or more. I would predict that every conceivable relationship between intelligence and bilingualism *could* obtain, and that our task is not so much the determination of *whether* there is a relationship between the two but of *when* (i.e., in which socio-pedagogical contexts) *which kind* of relationship (positive, negative, strong, weak, independent or not) obtains. (P. 38)

To this I might add that in the end, all this research notwithstanding, the question of bilingualism and intelligence, of whether they are linked positively or negatively, will evaporate in the face of deeper issues surrounding both bilingualism and intelligence. The fundamental question is misguided, for it entails two key simplifying assumptions. The first assumption is that the effect of bilingualism—indeed, the human mind—can be reduced to a single dimension (ranging from "good" to "bad"), and that the treatment (bilingualism) moves the individual child's standing up or down the dimension. The second assumption is that choosing whether the child is to be raised

## MIRROR OF LANGUAGE

bilingually or not is like choosing a brand of diaper, that it is relatively free of the social circumstances surrounding the choice.

"Is it good or bad for children to be bilingual?" As these simplifying assumptions are scrutinized over the course of the remaining chapters of this book, the need to ask such questions should diminish proportionately.