Lessons from a Paradoxical Hypothesis: A Methodological Critique of the Threshold Hypothesis

Mitsunori Takakuwa
Meiji Gakuin University

1. Introduction

Up to the 1960s, bilingualism was regarded as having a harmful effect on cognitive development. Early studies (e.g., Laurie, 1890; Saer, 1923; Smith, 1923) showed that bilingual children scored lower on measures of verbal intelligence than monolingual children, concluding that bilingualism hampers children's development of intelligence and leads children to psychological confusion. More recent studies on bilingualism (e.g., Ianco-Worrall, 1972; Peal and Lambert, 1962) supported the opposing view that bilingualism has a positive effect on children's cognitive development. For example, by studying 10-year-olds from French schools in Montreal, Canada, Peal and Lambert (1962) found that the bilingual group scored higher than the monolingual group on the measures of both verbal and nonverbal intelligence. This finding contradicted the results of the previous research that bilingual children were considered to be cognitively inferior to monolingual children.

Cummins (1976) attempts to provide an adequate explanation for the inconsistent findings from studies on the relationship between bilingualism and cognitive development. The hypothesis proposed by Cummins to account for the data available at that time is called the *threshold hypothesis*. The threshold hypothesis proposes that there may be a threshold level of bilingual proficiency that children must attain in order to gain advantages and avoid disadvantages in their cognitive development. Cummins (1976) identified two thresholds, the higher and lower threshold levels of bilingual proficiency:

> This raises the possibility that there may be not one but two thresholds. The attainment of the first threshold would be sufficient to avoid cognitive retardation but the attainment of a second, higher, level of bilingual competence might be necessary to lead to accelerated cognitive growth. (p. 24)

In other words, bilingual children must attain the lower threshold level in either language to avoid negative effects on their cognitive development. However, the attainment of the lower threshold level does not guarantee bilingual children cognitive advantages. Bilingual children must attain the higher threshold level in both languages to obtain positive effects on their cognitive development. The threshold hypothesis with two threshold levels is shown in Figure 1.

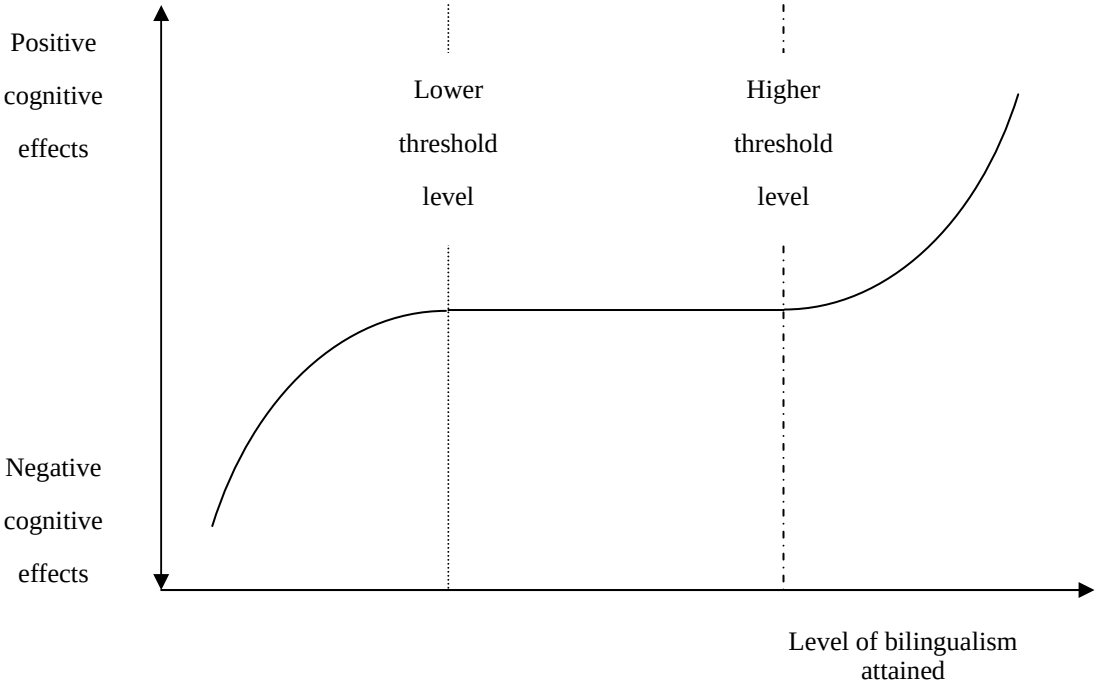
This hypothesis seems useful in accounting for the findings that some children benefit from bilingualism whereas others do not. According to the hypothesis, if bilingual children have not attained the lower threshold level in both languages, they are regarded as limited bilinguals and their language proficiency is at a low level in both languages. This situation is what Cummins (1979) calls *semilingualism*. If bilingual children attain the lower threshold level in either language, they are regarded as partial bilinguals and their language proficiency is at an age-appropriate level in only one language. Partial bilingualism has neither positive nor negative effects on children's cognitive development. This type of bilingualism is called *dominant bilingualism*. If bilingual children attain the higher threshold level in both languages, they are regarded as balanced bilinguals and their language proficiency is at an age-appropriate level in both languages. This type of bilingualism is called *additive bilingualism*. Thus, the findings of the early studies on bilingualism are associated with semilingualism whereas the results of more recent research are associated with additive bilingualism.

However, with some knowledge of methodology and common sense, it is easy to conclude that the threshold hypothesis is paradoxical and it is virtually impossible to support in a meaningful manner. If policymaking should be driven by tested theory, as Cummins (1999) himself asserts, theories in the research on bilingualism are integral to bilingual education policy. This means that incorrect conclusions based on a theory may affect policymaking in a way that students are put at a disadvantage. Although the hypothesis has already been criticized by other researchers (e.g., MacSwan, 2000; Martin-Jones & Romaine, 1986; Valadez, MacSwan, & Martinez, 2000), some researchers still claim that the hypothesis is supported by research (e.g., Cummins, 2000; Lasagabaster, 1998; Lee & Schallert, 1997; Ricciardelli, 1993; Roller, 1988; Schoonen, Hulstijn, & Bossers, 1998). Thus, additional discussion on the threshold

 ISB4: Proceedings of the 4th International Symposium on Bilingualism, ed. James Cohen, Kara T. McAlister, Kellie Rolstad, and Jeff MacSwan, 2222-2232. Somerville, MA: Cascadilla Press.

hypothesis is warranted.

Figure 1. Threshold Hypothesis Proposed by Cummins (1976)



This paper first examines critically the threshold hypothesis, and then addresses the methodological problems and limitations that are prevalent in the studies of bilingualism. This paper also discusses the importance of conducting research with a clear understanding of methodology.

2. The paradoxical nature of the threshold hypothesis

When addressing the paradoxical nature of the threshold hypothesis, it is useful to make a distinction between absolute and relative levels. Do the threshold levels proposed by Cummins (1976) exist, if any, in an absolute sense or in a relative sense? According to Cummins, threshold levels exist in the relative sense. He claims that "the threshold is likely to vary according to the type of bilingual learning situation and the individual's stage of cognitive development" (Cummins, 1976, p. 25). However, if Cummins' claim is valid and the threshold levels are relative, the threshold hypothesis is either meaningless or trivially true. The more studies claim to support the hypothesis, the more threshold levels are found (see Figure 2). This means that countless threshold levels are possible. It is even possible that the lower threshold level found in a study is higher, in the absolute sense, than the higher threshold level found in another (see Figure 3). When so many threshold levels exist, it may not make much sense to identify two of them arbitrarily and talk about them. For example, if the two threshold levels of an individual (or a class, a school, a district, etc.) are different from another, what can we conclude from the hypothesis in relation to bilingual education policymaking? A bilingual education program that takes into consideration the two threshold levels of an individual (or a class, a school, a district, etc.) may not accommodate those of another.

Figure 2. Threshold Hypothesis in the Relative Sense

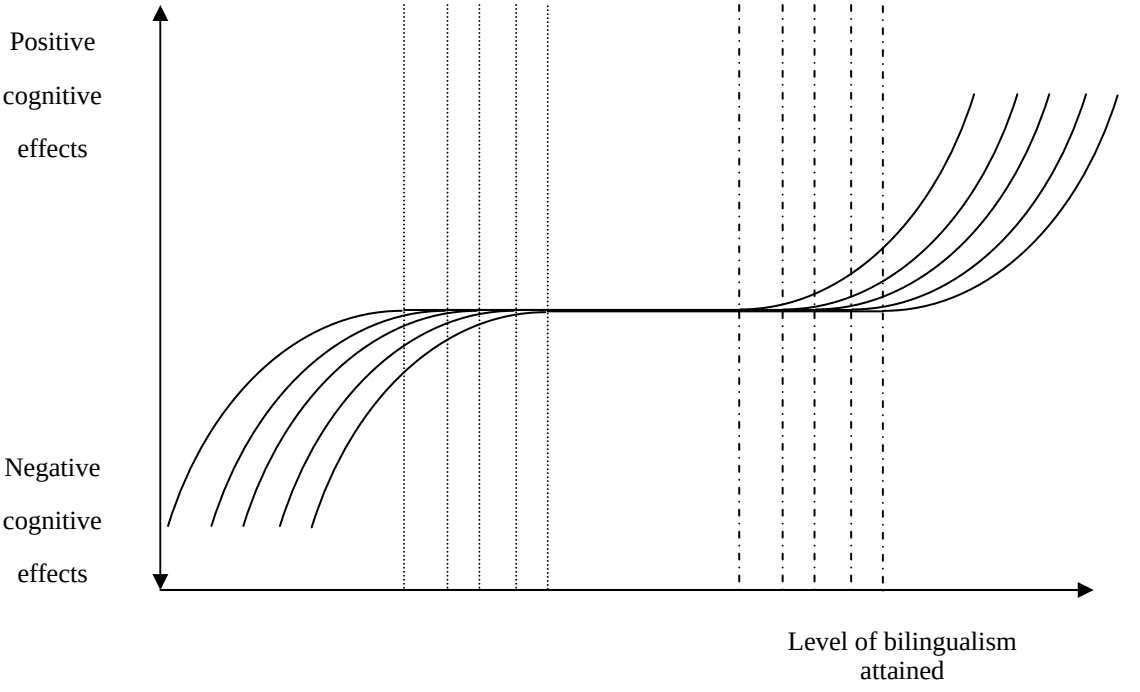
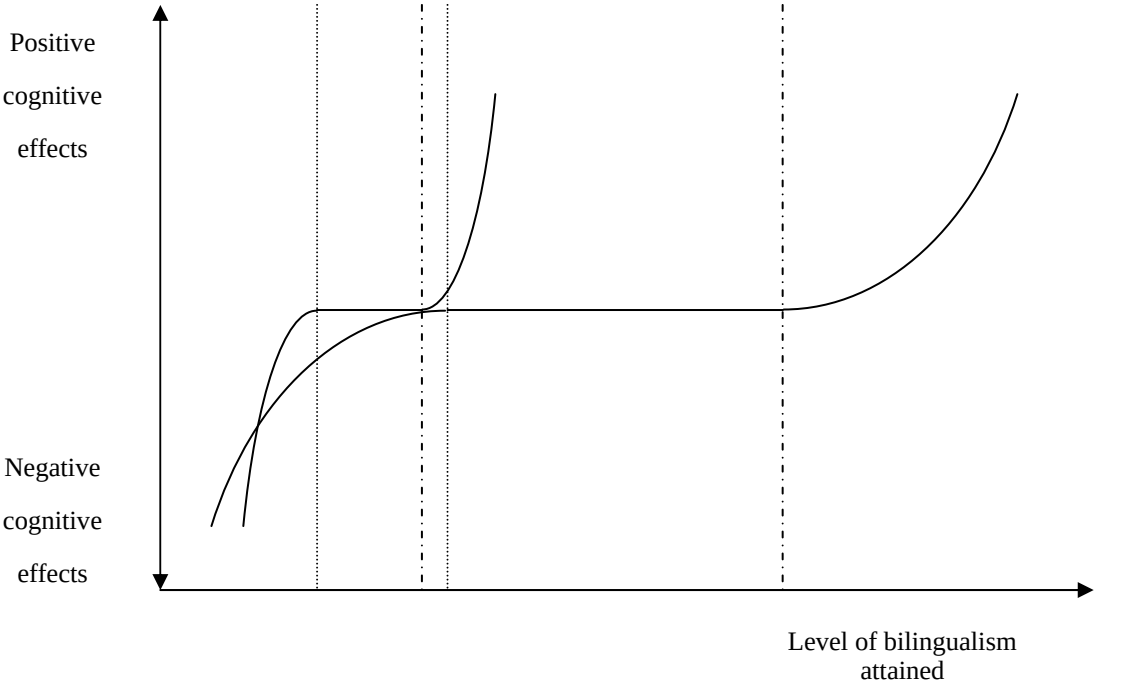
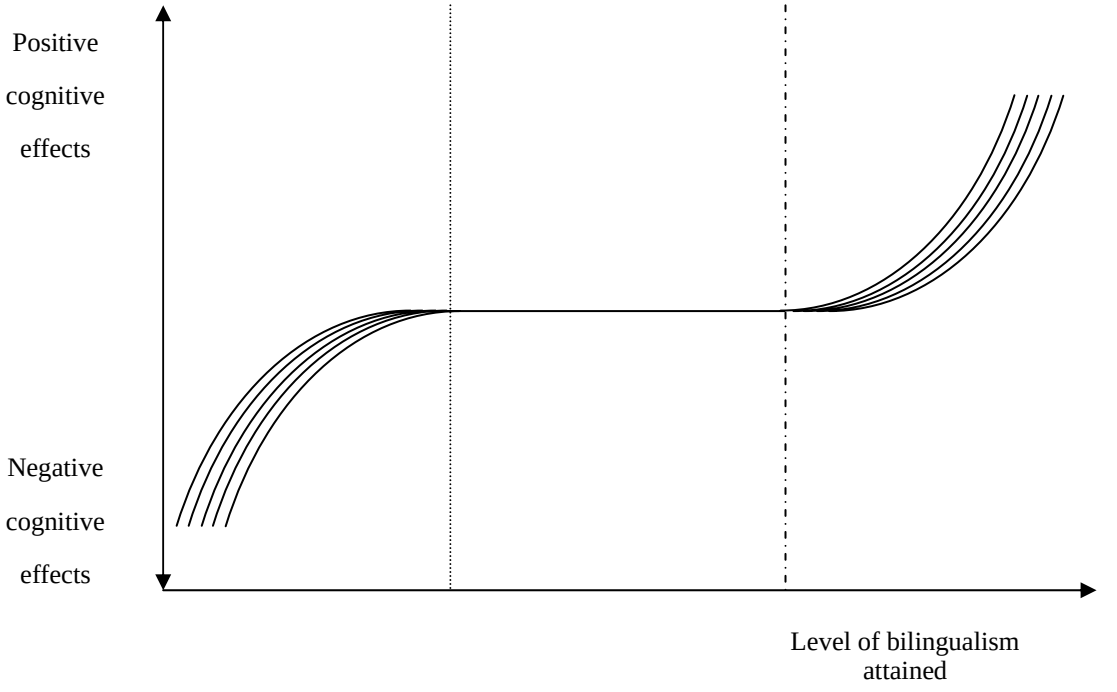


Figure 3. Threshold Hypothesis in the Relative Sense with the Two Levels Crossed



On the other hand, if the threshold levels are absolute, it is virtually impossible in a practical sense to conduct research that tests the threshold hypothesis. Testing the hypothesis in its absolute sense would involve administrating standardized tests on both L1 and L2 proficiency. If we could identify two particular points on the test scale as lower and higher thresholds across the studies, we would be able to say that there exist thresholds in the absolute sense and, therefore, the threshold hypothesis is tenable (see Figure 4).

Figure 4. Threshold Hypothesis in the Absolute Sense



However, no study has found such points and naturally the threshold hypothesis has not been supported in this way. This is not surprising given that there are very few, if any, standardized tests that are said to measure language proficiency and are standardized worldwide. For example, if one of the languages of bilingual students is English, TOEFL (Test of English as a Foreign Language) may be used to measure their English proficiency, but it is hard to find such tests to measure their proficiency in the other language, especially when there are not many people who use it as their native language. Therefore, the hypothesis is paradoxical. That is, in its relative sense, the more studies claim to support the hypothesis, the more threshold levels beyond the minimum are found. In its absolute sense, it is virtually impossible to conduct research that tests the threshold hypothesis and naturally no study has supported the hypothesis in this sense.

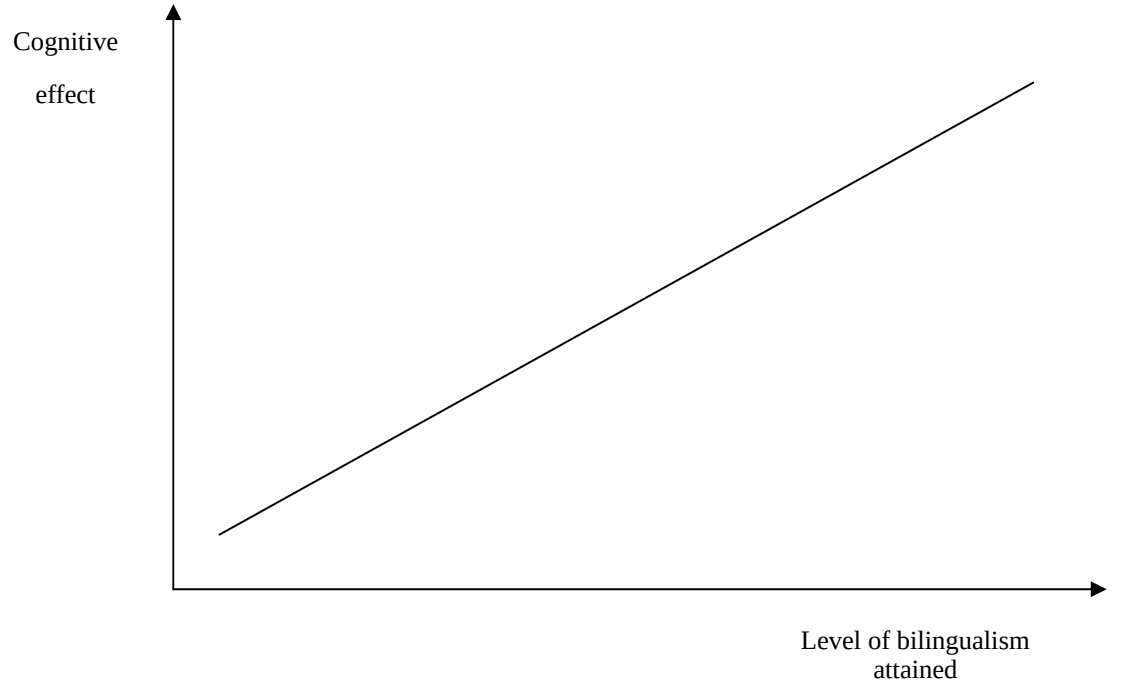
How, then, can some studies claim to support the threshold hypothesis? Although studies bearing on Cummins' threshold hypothesis display several methodological mistakes, misuse of categorization for continuous variables is the most serious problem. The important fact that is often neglected is that categorization for continuous variables is an arbitrary operation, and it establishes arbitrary levels where no such level exists in reality. It should be emphasized that the distinction between being bilingual and monolingual is a matter of *degree* rather than by an absolute discrete difference (Takakuwa, 2003). This has caused confusion in the literature regarding how to handle bilingualism as a variable. For example, the "monolingual" group in the Peal and Lambert (1962) study learned English at school as their second

language (L2), and understood it to the extent that they comprehended English words in "the test of English vocabulary which was administered by a native speaker of English" (p. 8). Saunders (1988) used the term monolingual, "for the sake of convenience, to refer also to persons near the extremity of the bilingualism continuum, namely to persons who are minimally bilingual, that is who have very little proficiency in more than one language" (p. 8).

Bilingualism is a matter of degree and it should be treated as a continuous variable. Similarly, one's cognitive development is a matter of degree and, therefore, the measures of one's cognitive development, whatever they are[1], should be treated as a continuous variable as well. Categorization of a continuous variable, such as dichotomization and trichotomization, is not only inappropriate but also uninformative. Such categorization adds errors of discreteness to the measurement error in the original scale (Cohen, 1983). Furthermore, if relationships among variables are linear, categorization reduces the degree of the relationship, and if relationships are nonlinear, categorization leads to a loss of information (Cohen, 1983; Humphreys & Fleishman, 1974; Pedhazur, 1997). Dichotomous distinctions on bilingualism and on cognitive development are not valid. Indeed, since such distinctions are arbitrary, one could create a nearly infinite number of subcategories.

The threshold hypothesis has been examined using categorization of continuous variables by means of the within-subjects criterion such as a median split or a mean split within their subjects (e.g., Lasagabaster, 1998; Ricciardelli, 1993). The problem that such a categorization process has is graphed

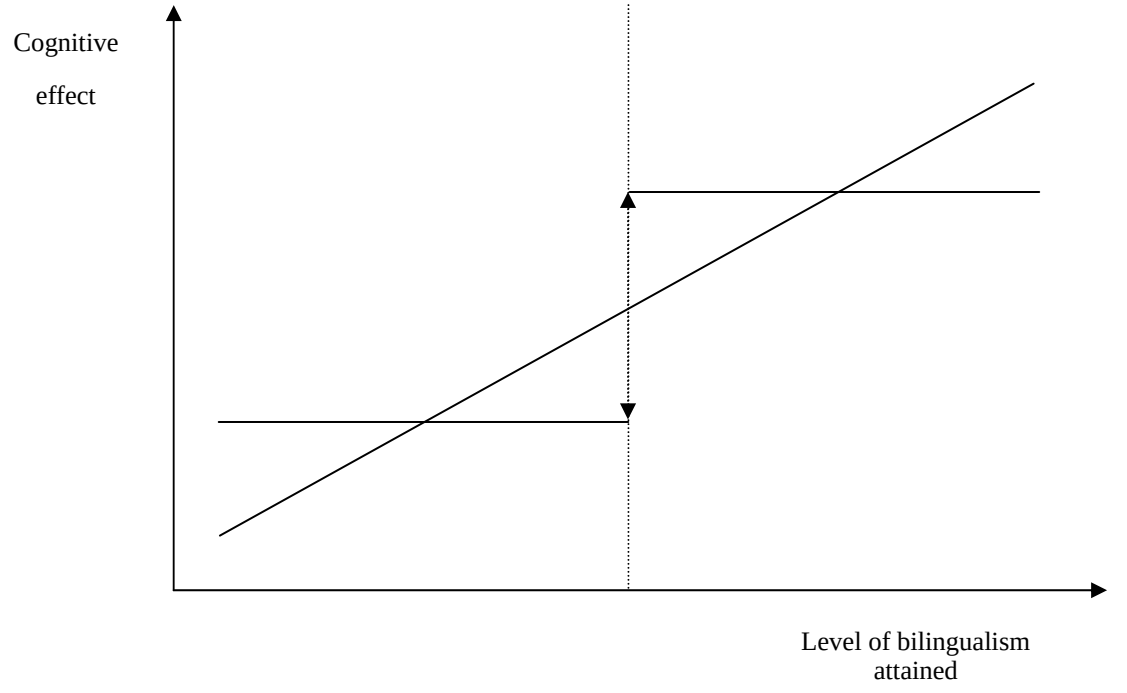
Figure 5. Original Relationship between the Level of Bilingualism and the Cognitive Effect



in Figures 5 and 6. When the relationship between the level of bilingualism attained and the cognitive effect is linear, the relationship is represented in Figure 5. However, if the categorization by means of a median split or a mean split is used, the result will be that there is a difference between the two arbitrarily formed groups which do not exist in reality, as the arrow in the middle of Figure 6 shows. Other studies (e.g., Roller, 1988; Schoonen, Hulstijn, & Bossers, 1998) have claimed to have found threshold levels between grades.

[1] The notion of cognitive development is vague and should be critically reviewed (Takakuwa, 2000). However, since the aim of this paper does not focus on what is meant by cognitive development, it will not be reviewed here.

Figure 6. Manipulated Relationship between the Level of Bilingualism and the Cognitive Effect



However, the choice of several grades itself is an arbitrary choice and it is far from clear if the differences among them would have been interpreted as threshold levels if the neighboring grades had been also included in the study. If a study arbitrarily establishes some threshold levels in and of itself, it is most likely to find them. This is the way some studies have claimed to support the threshold hypothesis.

3. Other lessons for research on bilingualism

It has been shown above that Cummins's threshold hypothesis is paradoxical and hard to support in a meaningful manner. This is largely due to misuse of categorization for continuous variables. However, the methodological problems in the studies on the threshold hypothesis are not limited to such categorization. Among the common mistakes frequently made in studies of bilingualism are overgeneralization, invalid use of tests of statistical significance, and inappropriate use of analysis of covariance. These methodological problems and limitations are prevalent in the field of bilingualism.

Overgeneralization has been a serious problem and yet has not been spotted in the literature. The schools (or districts, cities, etc.) in which studies on bilingualism were conducted were not randomly selected. This nonrandom sampling of the participants has a few important implications in conducting research. For example, the traditional distinction of samples and populations in statistical analyses necessarily involves random sampling (e.g., Glass & Hopkins, 1996; Tabachnik & Fidell, 1996). A population is any set of units or entities about which inferences are to be made based on a limited number of units at hand, called samples. To make valid inferences about a population, samples should be representative, and random sampling is a necessary (but not sufficient) condition for samples to be representative of a population. Glass and Hopkins (1996) emphasized this point and gave a warning about nonrandom sampling:

> Accidental or convenience sampling is a very common, but inappropriate, method of obtaining a sample. Convenient, but haphazard, collections of observations are usually of little value in

> estimating parameters. Results from street corner polls, polls of the audience of a particular television or radio program, or readers of a particular magazine cannot be generalized beyond such groups without great risks. (p. 226)

Thus, it is virtually impossible for samples that are selected nonrandomly to have a corresponding population to which valid inferences can be made based on those samples. This in turn means that the subjects selected nonrandomly are not samples of a population, but the population itself. Therefore, any results from a study without random sampling should not be generalized beyond the participants.

Another implication of nonrandom sampling is that hypothesis testing based on statistical significance is invalid. In educational and psychological research, the use of tests of statistical significance has been criticized for over 30 years, and the criticism by the methodologists led the American Psychological Association in the direction of reexamining the role of statistical significance testing (Wilkinson & APA Task Force on Statistical Influence, 1999). Even in a study with randomization, the use of tests of statistical significance alone is not recommended because statistical significance is, other things being equal, a function of the sample size. That is, a large enough sample can always make results statistically significant (see, e.g., Cohen, 1990, 1994; Kirk, 1996; Thompson, 1992, 1995, 1996).

Hypothesis testing based on statistical significance should be avoided with nonrandom sampling. For one thing, as discussed above, nonrandom samples do not have a population to which valid inferences can be made. However, what hypothesis testing based on statistical significance does is to test hypotheses about differences or effects in populations based on measurements made on samples. Thus, hypothesis testing based on statistical significance is *conceptually* invalid with nonrandom sampling. For another, hypothesis testing based on statistical significance is also *mathematically* invalid with nonrandom sampling. As Shaver (1993) put it:

> Randomness (i.e., random error) is the basis for the sampling distributions against which results are compared. Use of, for example, a *t* distribution to answer the question "How likely is this particular result under the null hypothesis?" will not yield a meaningful probability statement if the sample or samples are not random. Repeated random sampling (or assignment) yields known sampling distributions. Nonrandom sampling does not, nor does the comparison of a nonrandom sample to a randomly generated sampling distribution, provide a valid statement of probability of occurrence. (p. 295)

Thus, hypothesis testing based on statistical significance is not valid in studies with nonrandom sampling.

A further limitation of the studies on bilingualism is lack of random assignment. When comparing two or more groups, random assignment of the participants to each group is necessary. However, random sampling or random assignment is sometimes impossible in many studies of intact groups in education such as a class of students. The two or more groups in the studies could not be assigned randomly to treatment and control groups. When two or more intact groups have been compared in a study on bilingualism, it was often decided to employ a matching procedure and/or an analysis of covariance to reduce the irrelevant sources of variation as much as possible. A matching procedure was usually used in a way that two or more groups are matched on variables such as I.Q. and SES. However, a matching procedure reduces the initial differences of the groups only if matched subjects are randomly assigned to treatment and control groups (Glass & Hopkins, 1996). Using a matching procedure alone does not reduce the initial differences but rather potentially creates further problems because the selected variables used to match subjects are not necessarily appropriate or important (Borg & Gall, 1989; Rosenthal & Rosnow, 1991). Statistical control by means of analysis of covariance has been used in studies to reduce bias resulting from lack of randomization. It should be emphasized, however, that analysis of covariance will not equate any groups (Pedhazur, 1997) and, therefore, it will never be a substitute for randomization. That is, the initial group difference will not be fully eliminated with use of analysis of covariance. The initial group difference would be a confounding factor even if adjustment is taken into consideration in between-group analyses. Bias will always result from the lack of randomization. Therefore, caution is called for when interpreting the results.

The sample size of the studies should be discussed as well. The matter of sample size of a study is usually discussed in relation to hypothesis testing based on statistical significance (e.g., Cohen, 1990, 1994; Glass & Hopkins, 1996; Tabachnik & Fidell, 1996). In a study with randomization, other things being equal, if the sample size is too big, even a small difference or effect will be statistically significant, whereas if the sample size is too small, even a relatively large difference or effect will not be statistically significant. Thus, researchers of a study with randomization should be always careful of the sample size so that it may be neither too big nor too small. Power analysis may be used to determine the necessary

sample size to detect a population effect given a desired power, which is the probability of rejecting the null hypothesis when it is false, and a specified error rate (Cohen & Cohen, 1983).

In contrast, in a study without randomization, this type of discussion of the sample size in relation to hypothesis testing based on statistical significance is invalid because, as was seen above, hypothesis testing based on statistical significance itself is invalid without randomization. Rather, what is important about the sample size in a nonexperimental study concerns the matter of stability. In other words, if the sample size is too small, the relationship among variables can be erroneously large or small. For example, even among random numbers, for which the expected value of the correlation coefficient in terms of Pearson's *r* is $r = 0$, *r* can be larger than .7 if the sample size is N = 10. As the sample size gets bigger, the value of *r* is stabilized. When the sample size gets bigger than N = 150, the possible maximum value of *r* among random numbers will be about .2 or less. However, *r* can be larger than .4, for example, when N = 30, and it still can be larger than .2 when N=70 among random numbers. Thus, the relationship among the variables in the studies of bilingualism should be interpreted with caution.

4. The next step for the studies of bilingualism: toward more rigorous research

It has been shown above that Cummins's threshold hypothesis is paradoxical and hard to support in a meaningful manner. The methodological problems and limitations that are prevalent in the field of bilingualism research have also been pointed out. Precisely, Cummins may not have needed to propose the hypothesis. He tried to provide an adequate explanation for the inconsistent findings from studies on the relationship between bilingualism and cognitive development. However, it is not so clear if those findings are indeed inconsistent. Often the early studies have been criticized for failing to implement adequate methodological controls (e.g., Bialystok, 1988; Palij & Homel, 1987). Cummins also states that the early studies "suffered from methodological defects in that they failed to control for confounding variables such as socio-economic status (SES), sex and the degree of the bilingual's knowledge of his two languages" (Cummins, 1976, p. 3) and goes on to suggest a matching procedure by which bilinguals and monolinguals are matched on such variables. However, Cummins' view cannot go without criticism. That is, more recent studies on bilingualism, e.g., Peal and Lambert (1962) and subsequent studies, also suffer from methodological defects.

For example, Peal and Lambert's (1962) study is often considered to be a turning point in how bilingualism was viewed in the sense that, after their study, bilingualism became recognized as having a cognitive advantage (Palij & Homel, 1987). However, it is hard to support their findings as they claimed due to the incorrect labeling of the two groups in the study, the contradictory matching procedure for the two groups, and a blind faith in so-called intelligence tests. What Peal and Lambert really found was that, by comparing two groups in the same distribution of bilinguals rather than in the two different distributions of bilinguals and monolinguals (Reynolds, 1991), the one group at the higher end was superior to the other group at the lower end in terms of not only L2 proficiency but also first language (L1) proficiency, socioeconomic class, school grades, and so-called intelligence tests in which the test-taker's acquired language proficiency plays an important role (Miller, 1987; Oller, 1997).

Diaz (1985) reviewed research in the field of bilingualism and cognitive development and pointed out three methodological gaps that "seriously limit statements regarding the positive influence of bilingualism on children's cognitive development" (p. 1377). First, most studies used comparisons between bilingual and monolingual children. However, even if some variables such as age and SES are statistically controlled, "the matching procedure never guarantees that the two groups are equivalent in *all* relevant variables" (p. 1377). This between group comparison method has been also criticized by other researchers (e.g., Carey, 1991; Palij & Homel, 1987; Reynolds, 1991). Second, most studies focused on balanced bilinguals who have "native-like control of two languages" (Bloomfield, 1933, p. 56), not on nonbalanced bilinguals "who have disparate abilities in the two languages" (Diaz, 1985, p. 1377). However, there should be many students who are bilingual to some degree but are not completely balanced bilinguals, and the levels of bilingualism should be taken into consideration. Third, most studies investigated the relationship between bilingualism and cognitive development, but not the cause-effect relations between them. Hakuta and Diaz (1985) show that the direction of causality is, if anything, from bilingualism to cognitive development. However, Takakuwa (2000) argues that a different analysis of the same data shows that the direction of causality may be the opposite. Thus, it is unknown whether bilingualism has an effect, if any, on children's cognitive development or whether cognitively advanced children can become bilingual more easily. Even after two decades, Diaz's critique still holds for studies of bilingualism. Thus, studies utilizing a much more rigorous analyses are called for in the research on bilingualism.

5. Research on bilingualism as accumulation of scientific knowledge

To summarize, the methodological problems and limitations that are prevalent in the studies of bilingualism have been addressed in relation to the threshold hypothesis proposed by Cummins (1976). As we have seen, with a little knowledge in methodology and common sense, it is easy to conclude that the threshold hypothesis is paradoxical and it is virtually impossible to support in a meaningful manner. On the one hand, if the threshold levels are relative, the hypothesis is meaningless or trivially true. The more studies claim to support the hypothesis, the more threshold levels that are found. This means that the countless threshold levels are possible. If the threshold levels are absolute, on the other hand, it is virtually impossible to conduct research that tests the hypothesis. Therefore, the hypothesis is paradoxical.

This paper also discussed the importance of conducting research with a clear understanding of methodology. There have been studies that claim to support the paradoxical threshold hypothesis, and this may cause serious consequences in bilingual education. The most serious problem regarding research on Cummins' threshold hypothesis is the misuse of categorization for continuous variables. The important fact that is often neglected is that categorization for continuous variables is an arbitrary operation, and it establishes arbitrary levels where no such level exists in reality. The threshold hypothesis has been examined utilizing categorization for continuous variables by means of within-subjects criterion such as a median or mean split. If a study arbitrarily establishes some threshold levels itself, it is most likely to find them. Some other methodological problems found in the studies of bilingualism include overgeneralization, the use of tests of statistical significance, misuse of matching procedure and analyses of covariance, and sample size.

The rigorous approach to the analyses discussed in this paper should interest researchers who aim to conduct scientific research. In particular, researchers in the study of bilingualism will find the methodological rigor discussed in this paper necessary in their research. As often criticized (e.g., Carey, 1991; Palij & Homel, 1987; Reynolds, 1991) as well as critically reviewed in the present paper, many studies of bilingualism and bilingual education have been found to have deficiencies in their designs and analyses. The main defect common to such studies is a certain lack of rigor. Incorrect or inappropriate analyses found in those studies make their results uninterpretable. A rigorous approach to analyses should help move the study of bilingualism and bilingual education in the direction of more rigorous and scientific research. In summary, the rigorous approach provides researchers with a sense of how a study in bilingualism and bilingual education may be conducted with rigor. This can contribute to reinterpreting results of previous studies of bilingualism and bilingual education and will eventually yield the results from rigorous research that are interpretable and warranted.

Decisions on language planning are often made based on research findings (Baker, 2002; Cummins, 1999). What would the consequences be if such decisions were made based on findings of the studies that had gone wrong? Language planning of a country may strongly influence its bilingual education. If wrong decisions were made in language planning, many children would be at risk of being placed at an academic disadvantage at school. As MacSwan (2000) points out, if teachers form deficit beliefs about students, those inadequate beliefs may have an adverse effect on student achievement. Cummins (1999) himself states that rigorous criterion of validity should be imposed on any hypothesis. In light of a rigorous application of criterion of validity, it is time to abandon the threshold hypothesis because it is paradoxical and it is virtually impossible to support in a meaningful manner.

Ironically, Cummins (1976) has made a valid comment on research on bilingualism in the same paper as he proposed the threshold hypothesis:

> The search for consistent research results is based on a false premise – i.e. that there is but one single phenomenon or state called “bilingualism” which ought to influence the mental lives of all bilinguals in much the same way. . . In short, each bilingual learning situation is unique and it is impossible to generalize from one bilingual learning situation to another. (p. 11)

This comment is rigorously valid and compatible with what has been stated in this paper concerning the methodological problems and limitations of research on bilingualism. Lessons from the methodological problems concerning the threshold hypothesis are generalizable to some other areas in applied linguistics. Research findings with methodological problems are a threat to the accumulation of scientific knowledge. To conduct scientific research, understanding of methodology is critical.

References

Baker, Colin. (2002). Bilingual education. In Robert Kaplan (Ed.), *The Oxford handbook of applied linguistics* (pp. 229-242). Oxford: Oxford University Press.

Bialystok, Ellen. (1988). Levels of bilingualism and levels of linguistic awareness. *Developmental Psychology, 24*, 560-567.

Bloomfield, Leonard. (1933). *Language*. New York: Holt.

Borg, Walter R., & Gall, Meredith D. (1989). *Educational research: An introduction* (5th ed.). White Plains, NY: Longman.

Carey, Stephen. (1991). The culture of literacy in majority and minority language schools. *The Canadian Modern Language Review, 47*, 950-976.

Cohen, Jacob. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7*, 249-253.

Cohen, Jacob. (1990). Things I have learned (so far). *American Psychologist, 45*, 1304-1312.

Cohen, Jacob. (1994). The earth is round ($p < .05$). *American Psychologist, 49*, 997-1003.

Cohen, Jacob, & Cohen, Patricia. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cummins, James. (1976). The influence of bilingualism on cognitive growth: A synthesis of research findings and explanatory hypotheses. *Working Papers on Bilingualism, 9*, 1-43.

Cummins, James. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*, 222-251.

Cummins, Jim. (1999). Alternative paradigms in bilingual education research: Does theory have a place? *Educational Researcher, 28*(7), 26-32.

Cummins, Jim. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Clevedon, England: Multilingual Matters.

Diaz, Rafael M. (1985). Bilingual cognitive development: Addressing three gaps in current research. *Child Development, 56*, 1376-1388.

Glass, Gene V., & Hopkins, Kenneth D. (1996). *Statistical methods in education and psychology* (3rd ed.). Boston: Allyn & Bacon.

Humphreys, Lloyd G., & Fleishman, Allen. (1974). Pseudo-orthogonal and other analysis of variance designs involving individual-differences variables. *Journal of Educational Psychology, 66*, 464-472.

Ianco-Worrall, Anita D. (1972). Bilingualism cognitive development. *Child Development, 43*, 1390-1400.

Kirk, Roger E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56*, 746-759.

Lasagabaster, David. (1998). The threshold hypothesis applied to three languages in contact at school. *International Journal of Bilingual Education and Bilingualism, 1*, 119-133.

Laurie, Simon S. (1890). *Lectures on language and linguistic method in school*. Cambridge, England: Cambridge University Press.

Lee, Jeong-Won, & Schallert, Diane. (1997). The relative contribution of L2 language proficiency and L1 reading ability to L2 reading performance: A test of the threshold hypothesis in an EFL context. *TESOL Quarterly, 31*, 713-739.

MacSwan Jeff. (2000). The threshold hypothesis, semilingualism, and other contributions to a deficit view of linguistics minorities. *Hispanic Journal of Behavioral Sciences, 22*, 3-45.

Martin-Jones Marilyn, & Romaine, Suzanne. (1986). Semilingualism: A half-baked theory of communicative competence. *Applied Linguistics, 7*, 26-38.

Miller, Scott A. (1987). *Developmental research methods*. Englewood Cliffs, NJ: Prentice-Hall.

Oller, John W., Jr. (1997). Monoglottosis: What's wrong with the idea of the IQ meritocracy and its racy cousins? *Applied Linguistics, 18*, 467-507.

Palij, Michael, & Homel, Peter. (1987). The relationship of bilingualism to cognitive development: Historical, methodological and theoretical considerations. In Peter Homel, Michael Palij, & Doris Aaronson (Eds.), *Childhood bilingualism: Aspects of linguistic cognitive, and social development* (pp. 131-148). Hillsdale, NJ: Lawrence Erlbaum Associates.

Peal, Elizabeth, & Lambert, Wallace E. (1962). The relation of bilingualism to intelligence. *Psychological Monographs, 76*(27, Whole No. 546).

Pedhazur, Elazar J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). New York: Holt, Rinehart, and Winston.

Reynolds, Allan G. (1991). The cognitive consequences of bilingualism. In A. G. Reynolds (Ed.), *Bilingualism, multiculturalism, and second language learning: The McGill conference in honour of Wallace E. Lambert* (pp. 145-182). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ricciardelli, Lina A. (1993). An investigation of the cognitive development of Italian-English bilinguals and Italian monolinguals from Rome. *Journal of Multilingual and Multicultural development, 14*, 345-346.

Roller, Cathy M. (1988). Transfer of cognitive academic competence and L2 reading in a rural Zimbabwean primary

school. *TESOL Quarterly, 22*, 303-318.
Rosenthal, Robert, & Rosnow, Ralph. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.
Saer, D. J. (1923). The effects of bilingualism on intelligence. *British Journal of Psychology, 14*, 25-38.
Saunders, George. (1988). *Bilingual children: From birth to teens*. Clevedon, England: Multilingual Matters.
Schoonen, Rob, Hulstijn, Jan, & Bossers, Bart. (1998). Metacognitive and language-specific knowledge in native and foreign language reading comprehension: An empirical study among Dutch students in grades 6, 8 and 10. *Language Learning, 48*, 71-106.
Shaver, James P. (1993). What statistical significance testing is, and what it is not. *The Journal of Experimental Education, 61*, 293-316
Smith, Frank. (1923). Bilingualism and mental development. *British Journal of Psychology, 13*, 271-282.
Tabachnick, Barbara G., & Fidell, Linda S. (1996). *Using multivariate statistics* (3rd ed.). New York: HarperCollins.
Takakuwa, Mitsunori. (2000). What's wrong with the concept of cognitive development in studies of bilingualism? *Bilingual Review, 25*, 225-237.
Takakuwa, Mitsunori. (2003). Reassessing findings of causality in studies of bilingualism. *Meiji Gakuin Review, 692*, 51-73.
Thompson, Bruce. (1992). Two and one-half decades of leadership in measurement and evaluation. *Journal of Counseling and Development, 70*, 434-438.
Thompson, Bruce. (1995). Exploring the replicability of a study's results: Bootstrap statistics for the multivariate case. *Educational and Psychological Measurement, 55*, 84-94.
Thompson, Bruce. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher, 25*(2), 26-30.
Valadez, Concepcion, MacSwan, Jeff, & Martinez, Corinne. (2000). Toward a new view of low-achieving bilinguals: A study of linguistic competence in designated "semilinguals". *The Bilingual Review, 25*, 238-248.
Wilkinson, Leland, & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604.

ISB4: Proceedings of the 4th International Symposium on Bilingualism

edited by James Cohen, Kara T. McAlister, Kellie Rolstad, and Jeff MacSwan

Cascadilla Press Somerville, MA 2005

Copyright information

ISBN 978-1-57473-210-8 CD-ROM
ISBN 978-1-57473-107-1 library binding (5-volume set)

Ordering information

To order a copy of the proceedings, contact:

Cascadilla Press
P.O. Box 440355
Somerville, MA 02144, USA

phone: 1-617-776-2370
fax: 1-617-776-2271
sales@cascadilla.com
www.cascadilla.com

Web access and citation information

This paper is available from www.cascadilla.com/isb4.html and is identical to the version published by Cascadilla Press on CD-ROM and in library binding.