NORMES

The Big Picture in Bilingual Education: A Meta-analysis Corrected for Gersten's Coding Error

Kellie Rolstad
Arizona State University

Kate Mahoney
State University of New York

Gene V Glass
Arizona State University

Abstract
In light of a recent revelation that Gersten (1985) included erroneous information on one of two programs for English Language Learners (ELLs), the authors re-calculate results of their earlier meta-analysis of program effectiveness studies for ELLs in which Gersten's studies had behaved as outliers (Rolstad, Mahoney & Glass, 2005). The correction resulted in a change in mean effect size from .08 to .19 for all outcome measures, from -.06 to .14 for (English) reading, from .08 to .17 for (English) math, and from -.01 to .10 for all Transitional Bilingual Education (TBE) studies. The revelation of Gersten's coding error, and the inconsistency of Gersten's studies with other studies reviewed, increases confidence in the conclusion that an "investigator effect" suppresses results favoring TBE in these studies. Removing Gersten's studies from the meta-analysis renders an effect size of 0.17 for TBE, nearly as high as for Developmental Bilingual Education (DBE). The authors argue that the most informative result is the effect size reported for studies involving ELLs in both treatment and control groups, with an average effect size for TBE of 0.23. The new analysis therefore strengthens the conclusions previously reached in the authors' original research supporting TBE over English-only approaches, and DBE over TBE.

Introduction

Language minority education is at a peculiar point in its history. Within the last few years, clarity and consensus regarding the effectiveness of bilingual instruction has emerged in the scientific literature, while the political environment has become more hostile than at any time since the passage of the Bilingual Education Act of 1968.

In Rolstad, Mahoney, and Glass (2005) (hereafter, RMG), we reviewed narrative summaries and meta-analyses examining the effectiveness of bilingual education. The meta-analyses and more recent narrative summaries favored the conclusion that bilingual education is an effective approach to raising academic achievement for English Language Learners (ELLs), a conclusion also consistent with the work completed subsequently to RMG (August & Shanahan, 2006; Slavin & Cheung, 2003). A puzzling source of data for us was Gersten (1985), which behaved as an outlier in our analysis. In the present paper, we note recent revelations in Rossell and Kuder (2005) that Gersten miscoded program descriptions in his study, and we produce a new meta-analysis corrected for the coding error.[1]

Rolstad, Mahoney and Glass's (2005) Meta-analysis

RMG used a corpus of 17 studies that were conducted in the years following Willig's (1985) meta-analysis. Unlike previous studies, RMG provided comparisons not only for Transitional Bilingual Education (TBE) and English-only approaches, programs in which English acquisition is the primary goal, but also for Developmental Bilingual Education (DBE), programs that promote the development and maintenance of the first language as well as English. Furthermore, RMG included as many studies as possible in the meta-analysis, without applying selection criteria bearing on study quality, as intended by the original developers of the method (Glass, 1976; Glass, McGaw, & Smith, 1981).

As an additional methodological contribution, RMG coded program models according to the descriptions provided in the studies rather than the labels themselves, as many studies were found to use program labels adopted by schools but which did not fit conventional definitions. RMG coded programs whose descriptions were more aligned with the conventional definition of

TBE as TBE, those more aligned with the conventional definition of DBE as DBE, and those more aligned with the conventional definition of an English-only program as EO. See Crawford (2004) for conventional definitions and discussion of program models.

RMG showed that TBE was consistently superior to all-English approaches, and that DBE programs were superior to TBE programs. In an analysis controlling for ELL status, RMG found a positive effect for bilingual education of .23 standard deviations, with outcome measures in the native language showing a positive effect of .86 standard deviations. Note that in the Appendix (the table originally published in RMG) Gersten's three average effect sizes contributed negatively to the meta-analysis. More specifically and by individual effect size, Gersten (1985) contributed three negative effect sizes; Gersten, Woodward, and Schneider (1992) contributed ten negative and two positive effect sizes; and Gersten and Woodward (1995) contributed eleven negative effect sizes. For further details, please see RMG.

Gersten's Coding Error

Gersten (1985) had reported that a larger percentage of children enrolled in a structured immersion program (75%) scored at or above grade level on standardized tests than children in a bilingual program (19%) at the end of second grade. The term "structured immersion" was derived from established Canadian models of French immersion where instruction is in the immersion language, but teachers are bilingual, trained in immersion methods, and use a specially-designed curriculum in a six-year-minimum program (Baker & deKanter, 1983).[2] Gersten (1985) does not present a description of his comparison group apart from labeling it "the district's bilingual program" (p. 189). However, because Gersten has written extensively on bilingual education, consistently expressing a preference for direct instruction in Structured Immersion (SI) over bilingual methods, we included the 1985 study in our analysis along with two other Gersten contributions, even though it lacked an actual definition or description of the bilingual education program. In RMG, we coded Gersten's SI program as an EO program and what Gersten called TBE was coded as TBE. However, Gersten revealed that he "now agrees that the district undoubtedly mislabeled their ESL program as a bilingual program" (as cited in Rossell & Kudor, 2005, p. 18, footnote 7), and that the

comparison was not between EO and TBE, as Gersten originally stated, but rather between SI and ESL Pullout.

Gersten's three articles contributed 26 individual effect sizes out of 67 (39% of the sample), which had a substantial influence on the mean effect size. We now have a better understanding of why one of the studies, Gersten (1985), differed so dramatically from the others in the meta-analysis; rather than comparing TBE with SI, as originally stated, it compared ESL Pullout and SI. Gersten's (1985) description of the SI program in his study depends on reference to general characteristics of SI outlined in Baker and de Kanter (1983).

> The key to a structured immersion is that all academic instruction takes place in English, but at a level understood by the students (Baker & de Kanter, 1983). At the same time, there are always bilingual instructors in the class who understand the children's native language and translate problematic words into the native language, answer questions phrased in the native language, help the children understand classroom routines, show them the bathrooms, lunchroom, and playground, and so forth. (p. 189)

Furthermore, Gersten (1985) indicated that bilingual aides were used in the SI program and delivered Spanish-language instruction in all academic subjects:

> The paraprofessional aides serve two major purposes in the program. They are trained (by the head teachers) to teach daily lessons to small groups of children in the reading and arithmetic programs. Essentially, they serve as additional teachers, allowing for small group instruction in all academic areas. In addition, the bilingual aides help the non-English speaking students adjust to the environment, occasionally serving as translators during a child's first few months. (p. 189)

Gersten (1985) appears to define SI as used in the study, then, as involving bilingual teachers and bilingual aides who provide academic content instruction in the children's native language. While no details are provided regarding the ESL Pullout program, such programs generally do not provide native language support of any kind (Crawford, 2004). Therefore, following the coding convention established in RMG, we regard Gersten's SI as more aligned with TBE, since it appears to have provided native language support,

and we take what Gersten has now revealed to have been ESL Pullout as a variety of EO.

A Recalculated Meta-analysis Corrected for Gersten's Coding Error

Recalculating the meta-analysis in the Table with the corrected coding for Gersten (1985), following these conventions, we see that the mean effect size for all outcome measures increases from .08 to .19, Reading (in English) increases from -.06 to .14, Math (in English) increases from .08 to .17, and all TBE studies increased from -.01 to .10. The revelation of a coding error for Gersten (1985), and the inconsistency of all three of Gersten's studies with the rest of the work we reviewed, increases our confidence that the "investigator effect" noted in RMG may justify removing all three of the Gersten studies. As shown in the Table, removing Gersten's studies renders an effect size for TBE of 0.17, nearly as high as for DBE. Because numerous factors other than language proficiency are known to contribute to lower academic achievement among ELLs (August & Hakuta, 1998; August & Shanahan, 2006), we argued in RMG that the most informative result is the effect size reported for studies involving ELLs in both treatment and control groups; as shown in the Table, the average effect size for TBE in these studies is 0.23, favoring bilingual approaches.

Conclusions

Meta-analysis is a useful tool for clarifying variation among studies reporting divergent findings. The original RMG analysis discovered curious effects associated with the Gersten studies, which behaved as outliers in the analysis. The coding error recently reported by Rossell and Kuder (2005) confirmed our suspicion, at least for Gersten (1985), that the results were incorrect. The new analysis reported in the Table strengthens the conclusions previously reached in RMG supporting TBE over English-only approaches, and DBE over TBE.

Table

Combining Effect Sizes by Grouping before and after Correcting for Gersten's Coding Error

	Before Correction			*After Correction*		
Grouping	*N* of ES	*M* ES	*SD* of ES	*N* of ES	*M* ES	*SD* of ES[1]
All outcome measures	67	0.08	0.67	67	0.19	0.65
Reading (in English)	16	-0.06	0.61	16	0.14	0.6
Math (in English)	15	0.08	0.42	15	0.17	0.39
All outcomes in native language	11	0.86	0.96	11	0.86	0.96
Without Gersten studies	58	0.17	0.64	58	0.17	0.64
All TBE studies	35	-0.01	0.45	32	0.1	0.24
All DBE studies	30	0.18	0.86	30	0.18	0.86
All studies comparing ELLs to ELLs	22	0.23	0.97	22	0.23	0.97

Note. [1]"*SD* of ES" is the standard deviation of the effect sizes.

References

August, D., & Hakuta, K. (1998). *Improving schooling for language-minority children: A research agenda*. Washington, DC: National Academy Press.

August, D., & Shanahan, T. (Eds.). (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum.

Baker, K., & de Kanter, A. (1983). Federal policy and the effectiveness of bilingual education (pp. 33-85). In K. Baker & A. de Kanter (Eds.), *Bilingual education*. Lexington, MA: DC Heath. Also published as Baker, K., & de Kanter, A. A. (1981). *Effectiveness of bilingual education: A review of the literature. Final draft report.* Washington, DC: Department of Education Office of Planning, Budget, and Evaluation.

Burnham-Massey, M. (1990). Effects of bilingual instruction on English academic achievement of LEP students. *Reading Improvement, 27*, 129-132.

Carlisle, R. S. (1989). The writing of Anglo and Hispanic elementary school students in bilingual, submersion, and regular programs. *SSLA*, 11, 257-280.

Carter, T. P., & Chatfield, M. L. (1986). Effective bilingual schools: Implications for policy and practice. *American Journal of Education, 95*(1), 200-232.

Crawford, J. (2004). *Educating English learners: Language diversity in the classroom.* Los Angeles, CA: Bilingual Educational Services, Inc.

de la Garza, J., & Medina, M. (1985). Academic achievement as influenced by bilingual instruction for Spanish-dominant Mexican American children. *Hispanic Journal of Behavioral Sciences, 7*(3), 247-259.

Gersten, R. (1985). Structured immersion for language minority students: Results of a longitudinal evaluation. *Educational Evaluation and Policy Analysis*, *7*, 187-196.

Gersten, R., Woodward, J. (1995). A longitudinal study of transitional and immersion bilingual education programs in one district. *Elementary School Journal*, *95*(3), 223-239.

Gersten, R., Woodward, J., & Schneider, S. (1992). *Bilingual immersion: A longitudinal evaluation of the El Paso program*. Washington, DC: READ Institute. (ERIC Document Reproduction Service No. ED389162)

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*, 3-8.

Glass, G. V, McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: SAGE Publications.

Krashen, S. (2005). A correction – 20 years later. Stephen Krashen's Mailing List. Retrieved February 25, 2008 from http://sdkrashen.com/pipermail/krashen_sdkrashen.com/2005-November/000332.html

Lindholm, K. J. (1991). Theoretical assumptions and empirical evidence for academic achievement in two languages. *Hispanic Journal of Behavioral Sciences*, *13*(1), 3-17.

Medina, M., Jr., & Escamilla, K. (1992). Evaluation of transitional and maintenance bilingual programs. *Urban Education*, *27*(3), 263-290.

Medina, M., Jr., Saldate, M., IV, & Mishra, S. (1985). The sustaining effects of bilingual instruction: A follow-up study. *Journal of Instructional Psychology*, *12*(3), 132-139.

Medrano, M. F. (1986). Evaluating the long-term effects of a bilingual education program: A study of Mexican students. *Journal of Educational Equity and Leadership, 6,* 129-138.

Medrano, M. F. (1988). The effects of bilingual education on reading and mathematics achievement: A longitudinal case study. *Equity and Excellence, 23*(4), 17-19.

Ramirez, J. D., Yuen, S. D., Ramey, D. R., Pasta, D. J., & Billings, D. (1991). *Final report: Longitudinal study of immersion strategy, early-exit and late-exit transitional bilingual education programs for language-minority children.* San Mateo, CA: Aguirre International. (ERIC Document Reproduction Service No. ED330216)

Rolstad, K. (2008). Immersion education. In J. Gonzalez (Ed.), *Encyclopedia of bilingual education* (pp. 259-264). Thousand Oaks, California: Sage Publishers.

Rolstad, K., Mahoney, K., & Glass, G. V. (2005). The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational Policy, 19*(4), 572-594.

Rossell, C. (1990). The effectiveness of educational alternatives for limited-English proficient children. In G. Imoff (Ed.), *Learning in two languages* (pp. 71-121). New Brunswik, NJ: Transatlantic Publishers.

Rossell, C., & Kuder, J. 2005. Meta-murky: A Rebuttal to recent meta-analyses of bilingual education. In J. Sohn (Ed.), *The Effectiveness of Bilingual School Programs for Immigrant Children* (pp. 43-76). Berlin: Arbeitsstelle Interkulturelle Konflikte und gesellschaftliche Integration.

Rothfarb, S., Ariza, M. J., & Urrutia, R. (1987). *Evaluation of the Bilingual Curriculum Content (BCC) Pilot Project: A three year study (Final report).* Miami, FL: Dade County Public Schools, Office of Educational Accountability. (ERIC Document Reproduction Service No. ED300382)

Saldate, M., IV, Mishra, S., & Medina, M., Jr. (1985). Bilingual instruction and academic achievement: A longitudinal study. *Journal of Instructional Psychology*, *12*(1), 24-30.

Slavin, R. E., & Cheung, A. (2003). *Effective reading programs for English Language Learners: A best-evidence synthesis.* Baltimore, MD: Johns Hopkins University Center for Research on the Education of Students Placed At Risk (CRESPAR).

Texas Education Agency. (1988). *Bilingual/ESL Education: Program evaluation report*. Austin, TX: Texas Education Agency. (ERIC Document Reproduction Service No. ED305821)

Willig, A. C. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research, 55*(3), 269-318.

Endnotes

1. We are indebted to Stephen Krashen for bringing this important fact to our attention (Krashen, 2005).

2. The term "structured English immersion" (SEI), mandated in California and Arizona, tends to be used independently of "structured immersion" and does not require teachers to be bilingual, to be trained in immersion methods, or to use a specially-designed curriculum; moreover, SEI is intended as a one-year program, unlike the six-year SI program (Rolstad, 2008).

Appendix

Comparisons of Effect Size by Study as They Appeared in Rolstad, Mahoney & Glass (2005)

Study	*N* of ES	*M* ES	*SD* of ES[1]
Burnham-Massey, 1990			
Grades 7-8			
n(range) for TBE: 36-115			
n(range) for EO[2]: 36-115			
TBE vs EO[2]			
Reading	3	-0.04	0.07
Mathematics	3	0.24	0.14
Language	3	0.16	0.25
Carlisle, 1989			
Grade 4, 6			
n for TBE:23			
n for EO[1]:19			
n for EO[2]:22			
TBE vs EO[1]			
Writing-Rhetorical Effectiveness	1	0.82	
Writing- Overall Quality	1	1.38	
Writing-Productivity	1	0.60	
Writing-Syntactic Maturity	1	1.06	
Writing-Error Frequency	1	0.50	
TBE vs EO[2]			
Writing-Rhetorical Effectiveness	1	-2.45	
Writing- Overall Quality	1	-8.25	
Writing-Productivity	1	0.18	
Writing-Syntactic Maturity	1	0.24	
Writing-Error Frequency	1	1.01	
Carter and Chatfield, 1986			
Grades 4-6			
n(range) for DBE: 26-33			
n(range) for EO[2]:14-47			
DBE vs EO[2]			
Reading	3	0.32	0.24
Mathematics	3	-0.27	1.06
Language	3	-0.60	1.54

Appendix (continued)

Study	*N* of ES	*M* ES	*SD* of ES
de la Garza and Medina, 1985			
Grades 1-3			
n(range) for TBE: 24-25			
n(range) for EO[2]: 116-118			
TBE vs EO[2]			
Reading Vocabulary	3	0.15	0.38
Reading Comprehension	3	0.17	0.06
Mathematics Computation	3	-0.02	0.15
Mathematics Concepts	3	-0.02	0.14
Gersten, 1985			
Grade 2			
n(range) for TBE: 7-9			
n(range) for ESL: 12-16			
TBE vs ESL			
Reading	1	-1.53	
Mathematics	1	-0.70	
Language	1	-1.44	
Gersten, Woodward, and Schneider, 1992			
Grades 4-6			
n(range) for TBE: 114-119			
n(range) for ESL: 109-114			
TBE vs ESL			
Reading	4	-0.17	0.12
Language	4	-0.35	0.26
Mathematics	4	0.00	0.17
Gersten and Woodward, 1995			
Grades 4-7			
n for TBE: 117			
n for ESL: 111			
TBE vs ESL			
Reading	4	-0.15	0.13
Language	4	-0.33	0.22
Vocabulary	3	-0.15	0.12

Appendix (continued)

Study	*N* of ES	*M* ES	*SD* of ES
Lindholm, 1991			
Grades 2-3			
n(range) for DBE: 18-34			
n(range) for EO[1]: 20-21			
DBE vs EO[1]			
Reading	1	-0.59	
Language	2	-0.14	0.57
Medina and Escamilla, 1992			
Grades K-2			
n for DBE: 138			
n for TBE: 123			
DBE vs TBE			
language-oral, native	2	0.64	0.74
language-oral, English	1	0.11	
Medina, Saldate, and Mishra, 1985			
Grades 6, 8, and 12			
n for DBE:19			
n(range) for EO[1]: 24-25			
DBE vs EO[1]			
MAT Test			
Total Mathematics	2	-0.32	0.16
Problem Solving	2	-0.24	0.13
Concepts	2	-0.34	0.25
Computation	2	-0.13	0.53
Total Reading	2	-0.21	0.08
Reading	2	-0.30	0.28
Word Knowledge	2	-0.10	0.10
CAT Test			
Total Mathematics	1	-0.20	
Concepts/Application	1	-0.11	
Computation	1	-0.27	
Total Reading	1	-0.63	
Comprehension	1	-0.57	
Vocabulary	1	-0.41	

Appendix (continued)

Study	*N* of ES	*M* ES	*SD* of ES
Medrano, 1986			
Grades 1, 6			
n for TBE: 179			
n for EO[2]: 108			
TBE vs EO[2]			
Reading	2	-0.18	0.13
Mathematics	2	0.10	0.24
Medrano, 1988			
Grades 1, 3			
n for TBE: 172			
n for EO[2]: 102			
TBE vs EO[2]			
Reading	1	0.10	
Mathematics	1	0.60	
Ramirez, Yuen, Ramey, Pasta, and Billings, 1991			
Grades 1-3			
n(range) for DBE: 97-197			
n(range) for TBE:108-193			
n(range) for ESL: 81-226			
DBE vs ESL			
Mathematics	3	0.26	0.22
Language	3	-0.43	-0.97
Reading	3	0.37	0.21
TBE vs ESL			
Mathematics	3	0.11	0.10
Language	3	-0.17	0.17
Reading	3	0.01	0.16

Appendix (continued)

Study	*N* of ES	*M* ES	*SD* of ES
Rossell, 1990			
Grades K-12			
n for TBE: 250			
n for ESL: 326			
TBE vs ESL			
oral language	2	0.36	0.23
Rotharb and colleagues, 1987			
Grades 1-2			
n(range) for TBE: 34-70			
n(range) for ESL: 33-49			
TBE vs ESL			
Tests in English			
Mathematics	4	0.13	0.11
Language	2	0.28	
Social Studies	4	0.20	0.13
Science	4	0.09	0.18
Tests in Spanish			
Mathematics	4	0.11	0.14
Language	2	0.10	
Social Studies	4	0.23	0.22
Science	4	0.16	0.11
Saldate, Mishra, and Medina, 1985			
Grades 2-3			
n for DBE: 31			
n for EO[1]: 31			
DBE vs EO[1]			
Tests in English			
Total Achievement*	1	-0.29	
Reading	1	1.47	
Spelling	1	0.50	
Arithmetic	1	1.16	
Tests in Spanish			
Total Achievement	1	0.46	
Reading	1	2.31**	
Spelling	1	3.03	
Arithmetic	1	1.16	

Appendix (continued)

Study	*N* of ES	*M* ES	*SD* of ES
Texas Education Agency, 1988			
Grades 1, 3, 5, 7, 9			
n for TBE: approximately 135,000			
n for ESL: approximately 135,000			
TBE vs ESL			
Tests in English			
Mathematics	4	-0.03	0.02
Reading	4	-0.06	0.13
Tests in Spanish			
Mathematics	2	0.33	0.06
Reading	2	0.78	0.09

Note. *Reading, Spelling, and Arithmetic are not constituents of the Total Achievement; **This effect size was calculated with the treatment group's standard deviation; TBE is Transitional Bilingual Education; DBE is Developmental Bilingual Education; ESL is English as a Second Language; EO^1 is English Only instruction for Limited English Proficient children; EO^2 is English Only instruction for non-Limited English Proficient children. Permission to reprint material was obtained from *Educational Policy.*

Mexicans in the Pacific Northwest: Lessons from Progressive School Leaders for Progressive Educational Policy

Sheila M. Shannon
University of Colorado at Denver

Abstract
Latinos now live and work in areas of the United States where they have not been before. These changes impact schools in a variety ways. This article reviews recent research on how communities have responded in the South, New England and the West with a primarily assimilationist approach including English-only policies. The article then provides a description of one school district's response in the Pacific Northwest. This school district's progressive leadership provides guidance for progressive educational policy. The author concludes with a recommendation that the Bilingual Education Act of 1968 that No Child Left Behind eclipsed be reinstated to guide the nation in these changing times.

Introduction

With the passage of No Child Left Behind in 2002, bilingual education was reduced to lower case letters.[1] It no longer benefits from regular attention at the federal level, as it was when the Bilingual Education Act of 1968 was reauthorized with each new administration. As the current administration launches educational initiatives and mandates, it continues to ignore the dramatic demographic changes that the country it serves is experiencing. These changes are impacting schools in unprecedented ways and posing enormous challenges to schools. No other demographic changes have made this impact more than those associated with immigration primarily from Mexico and the New Latino Diaspora (Murrillo & Villenas, 1997).

Contemporarily, the Mexican diaspora has been "one *sui generis* whose dispersion to other countries is limited to [the United States of America]" (González Gutiérrez, 1999). The "New Latino Diaspora" refers to the migration