

Closing the Literacy Achievement Gap with Early Intervention

Emily M. Rodgers, Chuang Wang and Francisco X. Gómez-Bellengé

The Ohio State University

Paper presented at the 2004 Annual Meeting of the American Educational Research Association





lines. Achievement gaps are not limited to the United States; similar gaps have been documented between minority and majority cultures in other countries. In New Zealand for example, Maori and Pacific Island children typically achieve at lower levels than other children, and children in less economically advantaged schools achieve at lower levels than children in schools that are more economically advantaged (McNaughton, Phillips, & MacDonald, 2003).

It is not known why an achievement gap exists but race itself is certainly not a determining

gap differences of between 27 and 29 percentage points in fourth grade reading performance when comparisons are made across racial/ethnic or economic status lines. In fact, the state's annual report card for 2002 contains the disturbing observation that the gap is clear and consistent (Ohio Department of Education 2002).

<B> Early intervention and one-to-one teaching

One-to-one teaching is a recognized form of intervention for children having extreme

1. Does a literacy achievement gap exist along race/ethnicity and economic lines within a random sample of first grade students?
2. Do students who have had an opportunity for a full treatment of Reading Recovery, whether successful or not, close the literacy achievement gap along race/ethnicity and economic lines with a random sample of first grade students?
3. Do students who have been successfully discontinued from Reading Recovery (a subset of the treatment group) close the literacy achievement gap along race/ethnicity and economic lines with a random sample of first grade students?

#### <A> Methods

Data were gathered for three groups of first grade students on three literacy measures from *Assessment of Student Learning Achievement* (OS) (Clay, 2002) at two points in time: in the fall and spring of first grade (See Table 1). Results were disaggregated along two lines: race/ethnicity and economic status. These lines were selected because, as discussed in the literature review, research indicates that this is where the achievement gap lies: between races and also between more and less economically advantaged students (Lee, 2002).

#### <B> Data Sources

Data were gathered and analyzed for each student on three literacy tasks of the OS, a standard measure developed in research studies with established reliabilities and validities indices (Clay, 2002). The tasks are described in Table 1.

[Insert Table 1 about here]

Data were collected from three groups of students who were first graders in a Midwestern state during the 2002-2003 school year.

- 1) The treatment group. These were Reading Recovery students who had an opportunity to receive a full treatment of 20 lessons, whether they successfully completed the intervention (referred to as “discontinued”) or not (n= 4,764).
- 2) The discontinued group. This is a subset of the treatment group and includes only those children who met the criteria for successfully exiting the intervention because they were reading at average reading levels with their peers (n=3,499).
- 3) The comparison group. This group consisted of two randomly selected first grade children from each school in the state that had Reading Recovery (n=1,038).

Each group was disaggregated along race/ethnicity (African American or White) and economic status lines (measured by the student’s lunch cost status: regular or free). Of all the groups, the African American children in the random sample constituted the smallest size (n=126). As a result, samples of 126 children were randomly selected from each study group in order to have similar-sized groups for comparison purposes.

#### <B> Data Analysis

To determine whether a gap existed within the random sample along race/ethnicity and economic lines, we disaggregated data for each of the three measures and compared results for three groups (See Appendix A for each group’s mean scores):

1. African American random sample and White random sample (AARS & WRS)
2. Free lunch random sample and regular lunch random sample (FLRS & RLRS)

To compare the progress of all Reading Recovery students who had an opportunity to receive a full treatment with the progress of students in the random sample, we disaggregated data and compared results for these groups:

3. African American treatment group and White random sample (AATG & WRS)
4. Free lunch treatment group and regular lunch random sample (FLTG & RLRS)

To compare the progress of only those Reading Recovery students who were successfully discontinued with that of students in the random sample, we disaggregated data and compared results for these groups:

5. African American discontinued and White random sample (AADis & WRS)
6. Free lunch discontinued and regular lunch random sample (FLDis & RLRS)

Independent t-tests were conducted to determine if fall and spring gaps for the three reading measures existed between relevant groups on each measure. The alpha level was pre-set at .05 and an effect size estimate for each significant difference was calculated using Cohen's d with the pooled standard deviation (Rosnow & Rosenthal, 1996). In general, effect sizes of .25 and below are considered modest, those from .25 to .50 are moderate, and those above .50 are large (Cohen, 1977).

When there were significant differences in the spring scores, we did an additional analysis to measure mean gains from fall to spring. This analysis allowed us to assess the pedagogical significance of differences as well as to determine if the trend was towards opening, maintaining or closing the gap. (Descriptive statistics for fall-spring mean gains along race/ethnicity and economic status lines are contained in Appendices B and C respectively.)

#### <A> Findings

We first present findings that document the gap within the random sample along race/ethnicity and economic lines. Next, we describe the progress of the full treatment group towards closing the gap and finally, we consider the trend for the group of children who discontinued successfully from Reading Recovery. T-test and effect-size values for the gaps along race/ethnicity lines are reported in Table 2 and for the economic status line in Table 3.

[Insert Table 2 about here]

[Insert Table 3 about here]

<B> An achievement gap exists within the random sample



In the fall, even though mean scores for the WRS were higher than the AARS group on all three measures, the only significant difference existed for CAP ( $p < .05$ ). The 95% confidence interval for this comparison was from -2.713 to -.871 ( $d_{\text{FCAP}} = .48$ ). By spring, however, all of the differences between these two groups on the three measures were significant ( $ps < .05$ ;  $d_{\text{SCAP}} = .37$ ;  $d_{\text{SHRSW}} = .26$ ;  $d_{\text{STRL}} = .32$ ). These results suggest an opening of the gap along the race/ethnicity line within the random sample.

When we disaggregated the random sample data along the economic line we found that the means for the RLRS were higher than the FLRS on all three measures in both fall and spring. These differences were significant at the .05 level for all three measures at both points in time ( $d_{\text{FCAP}} = .70$ ;  $d_{\text{FHRSW}} = .70$ ;  $d_{\text{FTRL}} = .57$ ;  $d_{\text{SCAP}} = .46$ ;  $d_{\text{SHRSW}} = .40$ ;  $d_{\text{STRL}} = .69$ ). We interpret this to suggest that a significant gap existed in the fall along the economic line, and it remained opened at year end.

Having established a gap by end of year within the comparison group of first grade students on both disaggregated lines, we next compared the progress of the treatment group (all students who received a full treatment, regardless of outcome) to the random sample to determine if they closed the gap.

#### <B> A closing gap for the treatment group

When data were disaggregated along the race/ethnicity line for the treatment group (all students who had an opportunity to receive a full 20 weeks of lessons, whether successful or not), we found significant differences between the AATG and the WRS in the fall ( $ps < .05$ ) for all measures ( $d_{\text{FCAP}} = 1.16$ ;  $d_{\text{FHRSW}} = 1.16$ ;  $d_{\text{FTRL}} = .82$ ). By spring, means for the three measures remained higher for the WRS and the differences were still significant ( $ps < .05$ ;  $d_{\text{SCAP}} = .63$ ;  $d_{\text{SHRSW}} = .38$ ;  $d_{\text{STRL}} = .72$ ) but effect sizes were considerably reduced from fall to spring.

We interpret these findings to suggest that a gap existed in the fall of first grade between the African American students who received the intervention and Whites in the random sample, and

that the gap remained open in the spring. We found, however, that the AATG achieved higher fall to spring gains than the WRS on the HRSW and CAP measures, and the effect sizes were reduced, suggesting the gap is tending to close.

Next, we disaggregated the treatment group by economic status and compared the progress of FLTG students to RLRS students. Again we found significant differences in their scores in the fall and again in the spring ( $p < .05$ ;  $d_{\text{FCAP}} = 1.29$ ;  $d_{\text{FHRSW}} = 1.46$ ;  $d_{\text{FTRL}} = 1.03$ ;  $d_{\text{SCAP}} = .34$ ;  $d_{\text{SHRSW}} = .37$ ;  $d_{\text{STRL}} = .81$ ). We again found, however, that the FLTG had higher gain scores from fall to spring on the CAP and HRSW measures, meaning the trend was towards closing the gap. This trend to close the gap is also indicated by the reduction of the effect sizes from fall to spring.

<B> A closed or closing gap for the discontinued group

In the last set of comparisons, we compared the students who were discontinued successfully with the White Random Sample (WRS). When we disaggregated the discontinued group by race/ethnicity, we found significant differences on all three measures in the fall between the AADis and the WRS ( $p < .05$ ;  $d_{\text{FCAP}} = 1.05$ ;  $d_{\text{FHRSW}} = .96$ ;  $d_{\text{FTRL}} = .79$ ). By spring, these differences were no longer significant on CAP or HRSW but they still were for the TRL measure ( $p < .05$ ;  $d_{\text{STRL}} = .63$ ). These results indicate the gap closed along the race/ethnicity line for two of the three measures.

We did a secondary analysis to compare fall-spring mean gains because the differences on the TRL measure were still significant in the spring and found that the size of the gap was reduced for the AADis group relative to the WRS group. This is also reflected by a reduction in the effect size, even though the difference was still statistically significant.

In the final set of analyses, we compared the Free Lunch Discontinued (FLDis) group to random sample students who received regular-priced school lunches (RLRS) ( $p < .05$ ;  $d_{\text{FCAP}} = 1.12$ ;  $d_{\text{FHRSW}} = 1.51$ ;  $d_{\text{FTRL}} = .97$ ). We found significant differences on all three measures in the fall when data for the successfully discontinued group were disaggregated along the economic line. By spring,

differences were no longer significant on the HRSW or CAP measures but the difference remained statistically significant on the TRL ( $p < .05$ ;  $d_{\text{TRL}} = .63$ ). Although the gains on TRL for these two groups were similar, the effect size was reduced considerably. This suggests a closing of the gap on CAP and HRSW and a tendency to close the gap for TRL.

#### <A> Discussion

We found evidence within our random sample to support Denton and West's findings (2001) that a literacy gap exists between children as early as their second year at school and that it exists along racial/ethnic and economic lines, just as others have found (cf. West, Denton & Germin-Hauskin, 2000). Bainbridge and Lasley (2002) contend that it will take a system-wide approach to make a difference to the achievement gap; and that quick fixes will not do. We wondered whether intervening early with one-to-one teaching, using the example of Reading Recovery, might be a complex enough response to make a difference. Our findings suggest that it is.

We were not surprised to consistently find significant differences in fall scores between both Reading Recovery groups (the entire treatment group and the discontinued group) and the random sample because, by definition, students who qualify for Reading Recovery are having the greatest difficulty learning to read. It is also not surprising that differences between the treatment groups and the random sample groups remained significant in the spring, because the treatment groups contained all Reading Recovery students, including those who were successfully discontinued and those who were not.

Even so, gain scores on the HRSW and CAP measures showed unexpected progress for these lowest achieving students when results were disaggregated by economic and race/ethnicity lines. Although the differences between all RR students and RS students were still statistically significant in the spring, particularly the text reading measure (TRL), the effect sizes for these differences were reduced considerably.

The discontinued group of students also made unexpected gains, closing the gap along economic and race/ethnicity lines on the HRSW and CAP measures. Although the difference for TRL was still statistically significant in spring, it was much smaller than in fall (as evidenced by smaller effect sizes) and no longer pedagogically meaningful.

An analysis of the TRL measure indicates that a statistically significant gap still exists in spring between the RR groups and the corresponding RS groups, however, children in the disaggregated random sample groups experienced an opening of the gap on this measure. Results for the RR groups counter the trends observed in the general population not served by RR; instead of falling further behind, they tend to close the gap.

The progress of the Reading Recovery students runs counter to the progress that might be expected of low achieving children. Juel's longitudinal research suggests that it is extremely difficult for low achieving children to change their rank within their cohort: Once low, they tend to remain low achieving. (Juel, 1988).

Our findings support Bainbridge and Lasley's hypothesis that it will take a systemic effort, and not a "one-shot workshop" or a "quick fix" to change the achievement gap. Reading Recovery teachers take year long training at the graduate level, with weekly class sessions focused on the teaching of children, at the core of the training. This sustained professional development effort is one of the features of Reading Recovery that accounts for the progress of children, along with the nature of the instruction and the fact that it is delivered in a one-to-one setting, according to an experimentally designed study by Pinnell, Lyons, DeFord, Bryk and Seltzer (1994). Teachers who are trained in Reading Recovery take part in ongoing professional development sessions following their training, so they continue to focus on teaching and learning after their training year. The sustained nature of the professional development of Reading Recovery, along with the in-depth, long-term nature of the training, qualifies it, we think, as a systemic effort.



children to average levels of reading in first grade, we just may be able to spoil predictions of failure

<A> References

Bainbridge, W. L., & Lasley, T. (2002). Demographics, diversity, and K-12 accountability: The challenge of closing the achievement gap. *Education*, 34, 422-437.

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.

Clay, M. M. (1993/2002). *A... Portsmouth, NH: Heinemann.*

Clay, M. M. (2001). *C... Portsmouth, NH: Heinemann.*

Denton, K., & West, J. (2002). *C... (ERIC Document Reproduction Service No. ED461438).*

Denton, K., West, J., & Walston, J. (2003). Reading: Young children's achievement and classroom experiences. Retrieved February 9, 2004, from the National Center for Education Statistics

Ohio Department of Education (2002). *C... : 2002*. Columbus: Author.

Pinnell, G. S., Lyons, C. A., DeFord, D. E., Bryk, A. S., & Seltzer, M. (1994). Comparing instructional models for the literacy education of high-risk first graders. *R... Q...*, 29, 8-39.

Rosnow, R.L., & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *P... M...*, 1, 331-340.

Shanahan, T., & Barr, R. (1995). Reading Recovery: An independent evaluation of the effects of an early instructional intervention for at-risk learners. *R... Q...*, 30(4), 958-996.

U.S. Department of Education. (2001). *N... : F... 2000*. Retrieved June 25, 2003, from <http://nces.ed.gov/nationsreportcard/data/ppt/1>

V2001)1(0002 -5.2(g c1ter]osiAl06 Tc0.009ns)1(M)1(ltzcanlon16.4M. )]TJ/F1 19(trieved J)ch Q)/(da)1.f1 Tf121F



Zill, N., & West, J. (2000). *E ... : A ... A ...*

...

*F ...*, 2000 (ERIC Document Reproduction Service No. ED448899).



## Appendix B

## Descriptive Statistics for Fall-Spring Mean Gains along Race/Ethnicity Lines

Groups	Tests	n	Mean	SD
African American	Concepts About Print	115	6.36	3.42
Random Sample	HRSW	115	13.47	8.78
	Text Reading Level	115	13.60	6.95
African American	Concepts About Print	91	9.14	3.64
Reading Recovery	HRSW	91	20.42	7.88
Discontinued	Text Reading Level	91	16.23	3.66
African American	Concepts About Print	106	8.49	4.15
Reading Recovery	HRSW	106	21.06	7.97
Treatment	Text Reading Level	106	14.86	6.20
White Random Sample	Concepts About Print	115	5.73	2.85
	HRSW	115	11.60	8.83
	Text Reading Level	115	16.24	5.71
White	Concepts About Print	93	8.72	3.11
Reading Recovery	HRSW	93	21.19	7.87
Discontinued	Text Reading Level	93	18.14	4.65
White	Concepts About Print	99	8.24	2.85
Reading Recovery	HRSW	99	22.12	7.02
Treatment	Text Reading Level	99	15.72	6.06

## Appendix C

## Descriptive Statistics for Fall-Spring Mean Gains along Economic Lines

Groups	Tests	n	Mean	SD
	Concepts About Print	106	6.37	3.38
	HRSW	106	16.44	9.14

Table 1

A O... S... M... L... P...

---

A O... S... (Clay, 2002)

Task	Nature of Task	Range of Scores & Reliability
Concepts About Print (CAP)	Examines the child's concepts or understandings about print.	0-37 Reliability: Cronbach's alpha = .78 <sup>(1)</sup>
Hearing and Recording Sounds in Words (HRSW)	Measure of phonemic awareness. Student writes a dictated sentence. Five equivalent forms of the test are available.	0-37 Reliability: Cronbach's alpha = .96 <sup>(2)</sup>
Text Reading Level (TRL)	Oral reading measure. Teacher records all oral reading behaviors and determines an instructional reading level. Strategic problem solving activities are all also evaluated.	0-30 Reliability: Cronbach's alpha = .83

Table 2. Gaps along race/ethnicity lines

	Fall			Spring		
	FCAP	FHRWS	FTRL	SCAP	SHRSW	STRL
AARS vs	p < .05	p > .05	p > .05	p < .05	p < .05	p < .05
WRS	d = .48			d = .37	d = .26	d = .32
AATG vs	p < .05	p < .05	p < .05	p < .05	p < .05	p < .05
WRS	d = 1.16	d = 1.16	d = .82	d = .63	d = .38	d = .72
AADis vs	p < .05	p < .05	p < .05	p > .05	p > .05	p < .05
WRS	d = 1.05	d = .96	d = .79			d = .63

Notes: d = Effect size; FCAP = Fall Concept About Print; FHRWS = Fall Hearing and Recording Sounds of Words; FTRL = Fall Text Reading Level; SCAP= Spring Concept About Print; SHRSW = Spring Hearing and Recording Sounds of Words; STRL = Spring Text Reading Level; AARS = African American Random Sample; WRS = White Random Sample; AATG = African American Treatment Group; AADis = African American Discontinued; WRS = White Random Sample.

Table 3. Gaps along economic lines

	Fall			Spring		
	FCAP	FHRWS	FTRL	SCAP	SHRSW	STRL
FLRS vs	p < .05	p < .05	p < .05	p < .05	p < .05	p < .05
RLRS	d = .70	d = .70	d = .57	d = .46	d = .40	d = .69
FLTG vs	p < .05	p < .05	p < .05	p < .05	p < .05	p < .05
RLRS	d = 1.29	d = 1.46	d = 1.03	d = .34	d = .37	d = .81
FLDis vs	p < .05	p < .05	p < .05	p > .05	p > .05	p < .05
RLRS	d = 1.12	d = 1.51	d = .97			d = .70

Notes: d = Effect size; FCAP = Fall Concept About Print; FHRWS = Fall Hearing and Recording Sounds of Words; FTRL = Fall Text Reading Level; SCAP= Spring Concept About Print; SHRSW = Spring Hearing and Recording Sounds of Words; STRL = Spring Text Reading Level; FLRS = Free Lunch Random Sample; RLRS = Regular Lunch Random Sample; FLTG = Free Lunch Treatment Group; FLDis = Free Lunch Discontinued; RLRS = Regular Lunch Random Sample.