Is Early Literacy Intervention Effective for English Language Learners? Evidence from Reading Recovery

Jane Ashdown, New York University
Ognjen Simic, New York University

Abstract

The literacy achievement of 25,601 first-grade students who received Reading Recovery tutoring services, from school year 1992-93 to 1997-98, is examined in order to evaluate the performance of children in this group who were English language learners. The children in the Reading Recovery Group were compared with a Random Sample Group of 18,363 first graders drawn from the classroom population of children not identified as needing assistance, and with a Comparison Group of 11,267 first-grade children who were in need of Reading Recovery but did not receive it because of a lack of resources. The results suggest that Reading Recovery is an effective intervention that narrows the reading achievement gap between native and non-native speakers. Because some school administrators and teachers appear to lack confidence in the potential for non-native speaking children to benefit from this literacy intervention, implications of these perceptions are discussed with respect to key principles of Reading Recovery's implementation in schools.

Educators, parents, and policy makers continue to debate the most effective instructional approaches necessary to provide a meaningful education to English language learners; that is, children who are learning to speak English as an additional language (Collier, 1992; Wilkinson, 1998). In addition, there is continuing concern about educational inequalities in academic achievement between language-minority students and native English speakers (Cummins, 1986) as schools serve increasing numbers of English language learners from diverse language contexts (Hornberger, 1992; Lucas, Henze, & Donato, 1990). The purpose of this study was to evaluate Reading Recovery as a supplemental literacy program for first graders, and to discuss whether this early intervention contributes to English language learners' capacity to reach native speaker norms for academic achievement, specifically in terms of reading. In other words, we were interested in investigating whether Reading Recovery is effective as an instructional intervention for English language learners and, thereby, contributes to reducing inequalities in academic achievement between native and non-native speakers educated in monolingual English classroom contexts.

Research on Literacy Instruction for English Language Learners

In addressing the question of whether Reading Recovery is effective for children who are learning English as an additional language, we reviewed research studies in the following areas:

- Evaluation studies of the effectiveness of classroom literacy instruction on the reading achievement of children who are English language learners.
- Evaluation studies of the effectiveness of Reading Recovery as an early intervention for all children as well as English language learners.

Research on Classroom Literacy Instruction

An examination of research addressing the effectiveness of classroom literacy instruction for English language learners reveals that the field is dominated by questions regarding the use of a language other than English for instructional purposes. In particular, researchers have compared the academic achievement of students with English as a second language who have received classroom instruction in a variety of first and second language settings.

Ramirez and colleagues (Ramirez, Yuen, Ramey, & Pasta, 1991), compared outcomes for students in the more typical bilingual program adopted by schools, that is, "early-exit" instruction involving part-day Spanish instruction in kindergarten through second grade, with two alternative programs. These alternatives were (a) "late-exit" bi-lingual programs with initial instruction in Spanish, followed by balanced (50%/50%) instruction in English and Spanish from kindergarten through sixth grade, and (b) "structured immersion" programs with instruction given only in English. The Ramirez study was a longitudinal evaluation that followed children in each program from grades one to three. There were 319 children in the early-exit program, 233 children in the structured immersion program, and 170 children in the late-exit program. An additional group of 154 students in the late-exit bi-lingual program continued in the study from fourth to sixth grade in order to capture the particular outcomes of this instructional design.

According to Collier (1992) the Ramirez study confirmed evidence from numerous other investigations examining long-term achievement of English language learners. Improved academic achievement in a second language is positively related to the support children receive for education in their first language. For example, in Ramirez et al. (1991) the children in all three programs did equally well at first grade on the Comprehensive Test of Basic Skills in reading and mathematics. However, by fourth grade there were strong differences in academic performance between cohorts; notably children in the late-exit bilingual program were making faster progress in both English reading and math than children in the early-exit and structured immersion cohorts.

A meta-analysis of research on literacy achievement for English language learners, included in the report by the National Research Council's Committee on the Prevention of Reading Difficulties (National Research Council, 1998), confirms the potential risks to sustained achievement levels when children experience initial literacy instruction in a second language:

> The accumulated wisdom of research in the field of bi-lingualism and literacy tends to converge on the conclusion that initial literacy instruction in a second language can be successful, that it carries with it a higher risk of reading problems and of lower ultimate literacy attainment than initial literacy instruction in a first language, and that this risk may

compound the risks associated with poverty, low levels of parental education, poor schooling, and other such factors. (p. 234)

Despite these findings, school systems are often faced with few instructional choices other than immersion in monolingual English classes for English language learners. Schools have to identify instructional approaches that foster effective literacy learning for all children, including English language learners speaking a variety of primary languages, such as Spanish, Chinese, Russian, Arabic and many others. Many investigations in the area of literacy acquisition have examined the instructional contexts that best support such learners.

For example, New Zealand has recently experienced net migration gains of peoples from the Pacific Islands and Asia who speak a variety of languages. Wilkinson (1998) reported on the New Zealand data from an international evaluation of educational achievement in 32 countries. These data revealed that despite the high literacy levels of many nine- and fourteen-year old New Zealand students, those whose home language was different from the language of school (i.e., English) were performing below native English speakers on comprehension and word recognition measures. Frequent assessment of students' reading and regular reading aloud by the teacher were instructional practices correlated with closing the achievement gap on both these measures.

In summary, there is strong evidence of the positive impact on reading achievement of initial literacy instruction being conducted in a child's native language. However, the above research also suggests that where native language literacy instruction is not available, instructional practices that best support the literacy achievement of English language learners must be identified if inequalities in reading achievement are to be reduced.

Research on Reading Recovery Instruction

Many school systems, wanting to address the needs of "at-risk" literacy learners including those children who speak languages other than English, have implemented Reading Recovery as an early intervention and prevention program (delivered in English) that supplements classroom literacy instruction during first grade. Skilled teachers, specifically trained for the purpose, provide daily, 30-minute lessons to those children identified as having serious literacy learning difficulties and are the lowest performing readers in the cohort. The aim of Reading Recovery is to ensure that children receiving this individual tutoring catch up as quickly as possible with their classmates, usually in 16 to 20 weeks, so they can continue to make progress in reading and writing in a variety of classroom instructional contexts without needing further special assistance.

Reading Recovery for all students. There have been many evaluations of Reading Recovery conducted by those implementing the program. Lyons (1998) reviews over ten years of data collected as part of a national design, demonstrating the effectiveness of the program. From 1985 to 1997, a total of 436,249 first grade children entered the program, of which 60% met the criteria for discontinuing; that is, they read at or above the average of their class by the end of first grade and were able to continue to improve in literacy learning without needing further intervention. Most of the remaining children made progress, but did not have enough time in the school year to complete their programs. These are impressive results, considering that all children enrolled in Reading Recovery were the lowest performing readers in their first grade cohort.

Other studies, including those conducted by independent evaluators, have reported similar favorable results. Shanahan and Barr (1995), in their independent evaluation of Reading Recovery, conclude that Reading Recovery attains its stated goal by bringing the children's learning up to that of their average-achieving peers. They report that many children leave the program with well developed reading strategies, including phonemic awareness and spelling knowledge. However, the researchers point to problems in reporting approaches that may inflate the learning gains of Reading Recovery children. Shanahan and Barr call for clearer specifications of success, the documenting of outcomes on all students receiving Reading Recovery, and more rigorous research studies.

Other researchers of Reading Recovery who were seeking to assess the program's effectiveness, have suggested developing predictive models that would identify the characteristics of students most likely to succeed in Reading Recovery. For example, such a model has been proposed by researchers driven by cost-efficiency considerations (Batelle Institute, 1995). Identifying children more likely to succeed, it is argued, would drive down costs. By avoiding children predicted to fail, Reading Recovery could serve more children, more quickly.

Such an approach is dismissed by Reading Recovery professionals for practical and ethical reasons. By admitting the lowest scoring students, it is countered, Reading Recovery is potentially more cost-effective, because a significant number of these children who succeed in Reading Recovery do not later become a burden to the system, in terms of costly supplemental services in higher grades. In addition, children who are not among the most needy are the ones who are more likely to "survive" without costly special services, and benefit from classroom instruction alone.

Reading Recovery for English language learners. In our experience, English language learners, as a group, are students vulnerable to cost efficiency considerations and may be regarded as less likely to succeed in Reading Recovery as a monolingual English literacy intervention. Until recently there have been few attempts to disaggregate the impact of Reading Recovery on the performance of children who are learning English as another language. However, a study conducted in England included evidence of success of English language learners in Reading Recovery (Hobsbaum, 1995). More recently, Neal and Kelly (1999) examined reading and writing success for two groups of bi-lingual children receiving either Reading Recovery, where instruction is delivered in English, or Descubriendo La Lectura, a reconstruction of Reading Recovery, where intervention instruction is delivered in Spanish while children are receiving classroom literacy instruction in Spanish. The results indicated that both populations of students made progress and reached average levels of classroom literacy performance.

Purpose of the Study

Where bi-lingual education is not available, schools are faced with the challenge of how to foster high levels of literacy achievement for English language learners effectively. Evidence of Reading Recovery's effectiveness encourages school districts concerned with improving literacy achievement to adopt this program as a "safety net" for low performing students. We presumed it would be valuable to add to evaluations of Reading Recovery's contribution to the literacy achievement of English language learners, and to examine the extent to which it represents an appropriate educational program for this group of students.

In particular we were interested in whether such a contribution closes the achievement gap typically observed between native and non-native English speakers. To understand the impact of Reading Recovery on the reading achievement of first graders who are English language learners, we sought to answer the following questions:

1. Are there differences in outcomes, rate of completion, and delivery of Reading Recovery as a literacy intervention for children who are English language learners, as compared to native English speakers?
2. Does Reading Recovery narrow the gap in reading achievement between English language learners and native English-speaking children in first grade?

The focus of our attention centered on distinctions in Reading Recovery services and program performance between native and non-native English speakers. This reflects our broad interest in how, as an early literacy intervention, Reading Recovery works for children who have varying levels of competence in the English language.

Method

Measures and Criteria for Evaluating Success

The data used in this study were drawn from the *Reading Recovery Data Sheet*, produced by the National Data Evaluation Center at The Ohio State University. This is a national questionnaire used to record reading and writing scores, demographic information, and other data on all children selected for Reading Recovery, as well as on a sample of children randomly drawn from the general first grade classroom population.

Children are selected for Reading Recovery based on their performance on six literacy assessment tasks included in *An Observation Survey of Early Literacy Achievement* (Clay, 1993a), which were administered by Reading Recovery teachers. The children selected for services are the lowest performing first-grade children, deemed most "at-risk" of literacy failure in regular education classrooms. Clay (1993a) reports on the satisfactory measurement characteristics of the observation survey tasks, which assess letter identification (LI), sight reading vocabulary (Ohio Word Test = WT), concepts about print (CAP), writing vocabulary (WV), the capacity to hear and record sounds in words (HRSIW), and performance in reading a graded set of previously unseen texts (Text Reading Level = TRL). These graded texts have been benchmarked for use nationally in Reading Recovery and range in difficulty from pre-primer through sixth grade, leveled from 1 to 30 for use in first grade. For example, successful reading of levels 16 to 18 indicates appropriate grade level performance for the end of first grade to the beginning of second grade. In administering *An Observation Survey of Early Literacy Achievement* (Clay, 1993a) to children who speak English as a second language, there is a minimum requirement that they understand teacher-given directions for the tasks.

Evaluating success in Reading Recovery is based on two sources of information. One source is the combined judgments of the child's Reading Recovery teacher and the classroom teacher that the child is reading at or above the average performance of classroom peers. These judgments are checked against a second source of information, that is, testing at exit from the program using all six tasks on *An Observation Survey of Early Literacy Achievement* (Clay, 1993a). The specific decision to "discontinue" the tutoring of an individual child therefore depends on several sources of information and is evaluated against the following two criteria:

1. The extent to which the child has developed a self-sustaining learning system so that he or she can benefit from classroom instruction without the need for further intervention.

2. Results from exit testing by an independent observer (i.e., a teacher other than the child's Reading Recovery teacher) that indicate the child is reading close to his/her average performing peers. Note that the group's average band is based on the observation survey performance of a classroom random sample (mean +/- .5 SD), which is used as an empirical frame of reference to evaluate this achievement at the end of the school year.

Classroom reading achievement varies widely from district to district. "Discontinued from tutoring" as a label, is a relative criterion represented by varying achievement levels in different schools within sites (districts or collections of districts) implementing Reading Recovery. A "self-extending learning system" as a criterion for exiting the program depends on the clinical judgment of a Reading Recovery teacher that the child's observed reading and writing behaviors are evidence of cognitive capacities to make further literacy learning gains without continued individual tutoring (Clay, 1991). This criterion of a self-extending learning system is intended to be universal across all participating districts. A consistent teacher-training model in Reading Recovery, and continued support to teachers, ensures adherence to this criterion.

Participants

The selection of children into Reading Recovery is a nationally uniform procedure driven by the principles of its original design (Clay, 1993b). The children included in this study were initially identified for Reading Recovery as being among the lowest 20% of their first grade in reading according to their classroom teacher's judgment. Administration of *An Observation Survey of Early Literacy Achievement* (Clay, 1993a) by the Reading Recovery teacher provided further information to select the lowest performing children in need of immediate literacy tutoring. The national evaluation design calls for the testing of Reading Recovery children at the beginning of the school year, at program entry and exit, and at year-end.

The *Reading Recovery Group* included in this study comprised all children served regardless of their program status — successful, appropriately referred for specialist services including special education, having incomplete programs, or moved away from the school. Our choice was to include *all* of these students in the study, even if their exposure to Reading Recovery was minimal (a couple of lessons), in order to avoid any ambiguity in the definition of the intervention, a problem that plagued some previous research on Reading Recovery (see Shanahan & Barr, 1995).

Not all children who are initially identified as needing Reading Recovery eventually receive services. The most needy children are served first. Of the remaining children, some make progress through regular classroom instruction during the year, and thus do not need services. Others remain "at-risk," but do not receive Reading Recovery due to lack of resources. All of these children comprised the Comparison Group for this study. The evaluation design implemented by sites affiliated with New York University expands on the national design by collecting data on this *Comparison Group*, which we treat as an approximate solution for a control group of "at-risk" students.

The remainder of the classroom population, i.e., children generally considered not at-risk, served as a basis from which a random sample was drawn for each Reading Recovery site, again under the uniform procedures. The *Random Sample Group* was drawn from approximately the top 80% of

students in Reading Recovery classrooms and was tested at the beginning and end of the school year, using *An Observation Survey of Early Literacy Achievement* (Clay, 1993a), in order to provide a benchmark for reading achievement in a Reading Recovery site.

From the total number of 55,875 students in the groups (Reading Recovery, Random Sample, and Comparison) in NYU-affiliated sites, 644 children for whom information on native language was not available were eliminated from the study. The remaining 55,231 children were identified as English native speakers ("English" = 45,303 children), fluent non-native speakers ("Fluent ESL" = 6,388 children), and non-native speakers with limited English proficiency ("LEP" = 3,540 children) based on the data collected through the national Reading Recovery questionnaire (see Table 1). Children were characterized as such either through the results of a language proficiency test, if such a test was given by a district, or through classroom teacher judgment. All of the children came from monolingual classrooms, where instruction was in English.

Of all English language learners in the study, Spanish was the native language for the majority of the limited English proficient students (54%), with Chinese spoken by 26%, and other languages by 19%. Again Spanish was the dominant native language for language learners who were fluent in English — 74% spoke Spanish, 6% spoke Chinese, and 20% spoke other languages.

Reading Recovery Sites

The database used in this study spans six years of Reading Recovery implementation (school year 1992-93 to school year 1997-98) at 37 Reading Recovery sites affiliated with New York University. These sites, which may be a single school district or a consortium of districts working together to implement Reading Recovery, represent a variety of educational environments, including urban, suburban, and rural settings. Districts also varied in the number of years of Reading Recovery implementation, the number of certified Reading Recovery teachers available relative to need for service in schools (i.e., level of coverage), and the level of their experience in Reading Recovery.

Data Analyses

The first research question, which concerned the outcomes, completion rates, and delivery of Reading Recovery, was answered by a comparison of the proportion of children of different language

Table 1. Native Language Composition of Study Samples

| | Native Language | | | |
| | English | Fluent ESL | LEP | Total |
Sample				
Reading Recovery				
Count	20863	2924	1814	25601
%	81.5	11.4	7.1	100
Comparison Group				
Count	8845	1427	995	11267
%	78.5	12.7	8.8	100
Random Sample				
Count	15595	2037	731	18363
%	84.9	11.1	4.0	100

Note. LEP = Low English Proficiency

backgrounds who were selected to receive Reading Recovery services, who completed full Reading Recovery instruction, and who were deemed successful in Reading Recovery. Pearson's Chi-square tests were used to report on the statistical significance of the differences between two groups of English language learners (LEP and fluent ESL) and native English speakers (English).

To answer our second question, whether Reading Recovery closes the literacy achievement gap between native-speakers and English language learners in first grade, proved a challenging task, considering that our data derive from a field implementation of Reading Recovery in a variety of educational settings. To search for differences we used analysis of variance, with *language* (English, Fluent ESL, and LEP) and *sample group* (Reading Recovery, Random Sample, and Comparison) as fixed factors; *Reading Recovery Site* as a randomly varying factor; and *Text Reading Level* as a dependent variable.

By including *Reading Recovery Site* as a random factor in our model, we took into account the similarity of students within sites, due to shared curriculum, educational policies, geography, and other features. Differences between the groups of students who share educational settings are all the more important when one considers the heterogeneous nature of school systems that implement Reading Recovery in the wider New York metropolitan area. Including this source of variation explicitly provided us with better estimates of error and, thereby, gave us more confidence in estimates of effects, which were of primary interest to us. Specifically, the interaction of *language* and *sample group* effects represents a direct test of the hypothesis that the differences in reading achievement between language groups are smaller for Reading Recovery students than they are for the other two groups of first-graders (Random Sample Group and Comparison Group).

Results

Analysis of Outcomes, Completion Rates, and Delivery of Reading Recovery

The first study question concerned the extent to which there was any evidence of differences in outcomes and completion rates between Reading Recovery children from the three language groups. In addition we questioned whether there was equity in the delivery of Reading Recovery to children regardless of their native language background.

Program outcomes. We initially analyzed whether there were differences in outcomes for Reading Recovery children who were English language learners, consisting of fluent and limited English proficient,

Table 2. Reading Recovery Program Success and Program Completion Rates for Three Language Groups

| Language | | Outcome | | Completion | |
		Successful	Not	Complete	Incomplete
English	Count	12975	7888	15756	5107
	%	62.2	37.8	75.5	24.5
Fluent ESL	Count	1938	986	2253	671
	%	66.3	33.7	77.1	22.9
LEP	Count	1120	694	1348	466
	%	61.7	38.3	74.3	25.7
Total	Count	16033	9568	19357	6244
	%	62.6	37.4	75.6	24.4

as compared to native speakers. Table 2 presents program success rates for children in three different language groups (English, Fluent ESL and LEP children) as well as rates of program completion.

The success rates in Table 1 are expressed as a percentage of all students served in Reading Recovery. Of the 25,601 children served in this six-year period, 16,033 (63%) successfully exited the program, while the remaining 9,568 (37%) children were not successful. Since these figures account for all children served in the program, the "not successful" group includes children who moved from the school, those who lacked opportunity to complete a full program before the end of the school year, and those who were recommended for other services, including special education.

Statistically significant differences (chi-square = 18.960, df = 2, $p < 0.0001$) in success rates were observed among the language groups. Fluent ESL children have a higher success rate (66.3%) than either native English speakers (62.2%) or LEP students (61.7%). However, it appears that the limited English proficient children were just as successful as their native English-speaking peers.

Completion of Reading Recovery. In addition to considering success rates for Reading Recovery children, we also examined the extent to which children from different language groups had an opportunity to receive at least sixty lessons (a "full program" definition established when the program was first implemented in the United States), regardless of whether they successfully exited the program or not. We were interested in whether all Reading Recovery children had an equal opportunity to be successful, by receiving a full Reading Recovery program, regardless of their language background and English proficiency. Analysis of the data demonstrated that language proficiency was not a factor impacting children's opportunities to complete the program. There were no significant differences in program completion rates (see Table 1) between the three language groups (Chi-square = 5.046, df = 2, p = 0.08). In addition to student mobility, referral to special services was the most frequent reason for exiting the Reading Recovery program before completion. The analysis of completion rates suggests that these factors (mobility and referral) did not differentially impact Reading Recovery students from these three language groups.

Program delivery. In order to examine the selection process for Reading Recovery, we analyzed the language composition of each study sample — Reading Recovery Group, Comparison Group and Random Sample Group. We observed that the Reading Recovery Group contained a disproportionate number of native English speakers with respect to the Comparison Group (81.5% vs. 78.5%). This difference is statistically significant, as indicated by Pearson's Chi-square test (chi-square = 50.3, df = 2, $p < .0001$). Both Fluent ESL students (12.7% vs. 11.4%) and, especially, LEP students (8.8% vs. 7.1%), were less likely to be served in Reading Recovery than their peers who are native English speakers (and, thus, became part of the Comparison Group).

This finding is intriguing, considering that Reading Recovery targets the lowest performing first graders. It was revealed by an analysis of the measures from *An Observation Survey of Early Literacy Achievement* (Clay, 1993a), taken at the beginning of the school year, that LEP students who were selected into Reading Recovery indeed had somewhat higher scores than the LEP students who were not selected. In contrast, between English speaking and Fluent ESL students, it was clear that students with the lowest observation survey scores were the ones selected. This pattern of results indicates that sometimes decision-making may have been influenced by factors other than literacy assessment and this will be discussed later.

Table 3. Year-End Observation Survey Scores

| | Mean Score | | | |
Sample	TRL	WV	OWT	HRSIW
Reading Recovery				
English	15.73	46.33	16.74	33.77
Fluent ESL	15.26	47.20	16.79	33.31
LEP	14.61	48.09	16.52	33.10
Comparison Group				
English	12.21	37.55	14.95	30.68
Fluent ESL	10.05	34.45	13.53	28.26
LEP	7.66	32.92	12.12	25.54
Random Sample				
English	19.78	47.00	17.97	34.22
Fluent ESL	16.16	43.43	16.90	32.25
LEP	11.72	38.19	14.94	29.34

Note. TRL = Text Reading Level; WV = Writing Vocabulary; OWT = Ohio Word Test; HRSIW = Hearing and Recording Sounds in Words Test; LEP = Low English Proficiency.

Analysis of Reading Achievement by Group

Consistent with the Reading Recovery evaluation design, four of the tasks from *An Observation Survey of Early Literacy Achievement* (Clay, 1993a), including text reading, writing vocabulary, word recognition, and hearing and recording sounds in words, were administered at the end of the school year to all Reading Recovery children, as well as to the Random Sample, and to the children who were initially diagnosed as "at-risk" but were not served in Reading Recovery (i.e., the Comparison Group).

Average year-end scores for these three sample groups (see Table 3) appear to support the hypothesis that Reading Recovery closes the reading achievement gap between native and non-native English speakers. On all four measures, smaller differences in reading and writing achievement associated with native language proficiency were evident for Reading Recovery children.

The Text Reading Level (TRL) task is by far the most comprehensive and clinically meaningful of the tasks. The TRL scores provided in Table 3 represent the difficulty level achieved by students on a series of previously unseen, graded text passages, read with at least 90% accuracy. In the context of classroom instruction these results indicate that LEP children who had received Reading Recovery services were reading texts with a difficulty level equivalent to a Grade 1 basal reader. In contrast, LEP students in the Random Sample, a group that had not been identified as needing supplemental tutoring, were reading at only the Primer level at the end of first grade.

Table 4. Tests of Between-Subjects Effects: Spring Text Reading Level

Source	F	df	Sig.
Language	34.2	(2,194.3)	.000
Sample	65.1	(2,153.1)	.000
Site	7.1	(36,106.6)	.000
Language X Sample	11.7	(4,1093.9)	.000
Language X Site	3.3	(70,158.7)	.000
Sample X Site	4.5	(72,239.7)	.000
Language X Sample X Site	1.5	(129,50670)	.001

Spring TRL Grade Equivalent

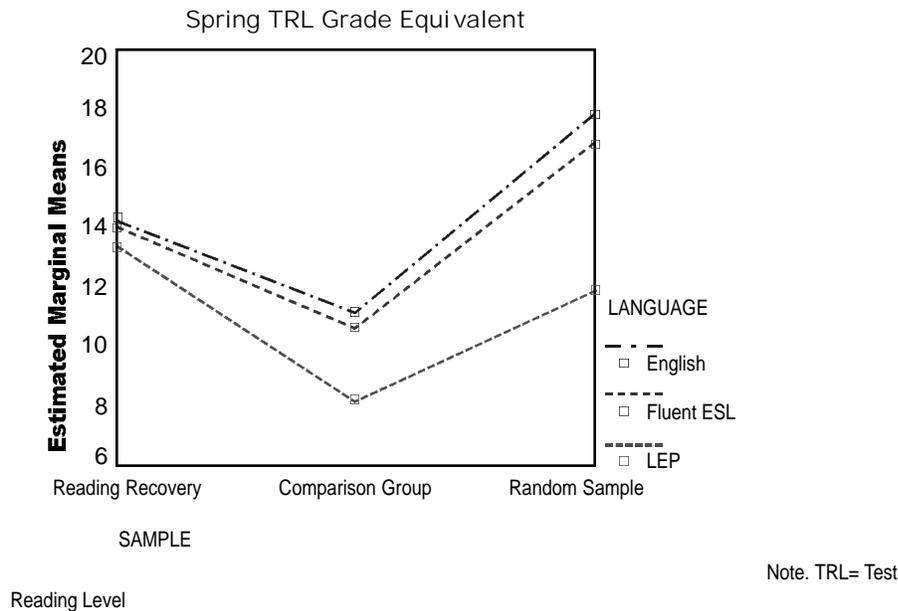

Note. TRL= Test
Reading Level

Figure 1. Interaction Language X Sample: Spring Text Reading Level

For two of these four observation survey measures, Hearing and Recording Sound in Words (sentence dictation task) and the Ohio Word Test (high-frequency word list), further statistical analysis was not advisable due to strong ceiling effects which resulted in skewed distributions of students' scores. For brevity's sake, we report the analysis of variance for Text Reading Level only, while noting that using Writing Vocabulary as a dependent variable led to exactly the same pattern of results.

An analysis of variance was conducted with *language* and *sample* as fixed factors, *site* as a random factor, and text reading level as the dependent variable. Tests of the main effects and interactions are presented in Table 4. The interaction of *sample* and *language*, which represents a direct test of the hypothesis regarding Reading Recovery's impact on the reading achievement gap, is graphically illustrated in Figure 1. All of the tests were statistically significant, and differences among means were in the expected direction.

It is apparent that the gap between the three language groups varied significantly, but was much smaller for children who received Reading Recovery, than for the children who did not. Non-native English-speaking children, especially LEP children, lagged behind native speakers both in the sample drawn from the lower (Comparison Group) and higher (Random Sample) end of the classroom reading achievement spectrum. Among Reading Recovery children these differences were drastically reduced.

On average, Random Sample children scored higher than those children considered "at-risk." This result inevitably follows from the evaluation design, where one group is sampled from the higher and the other from the lower end of the achievement range. Also, our definition of "Reading Recovery children" was all-inclusive, and did not omit children who were either unsuccessful or had incomplete programs. When the same analysis was conducted using only the children who were successful (63% of the cohort), the difference between Reading Recovery and Random Sample disappeared, as Reading

Recovery children scored on the level of their peers. The same pattern of results was evident when Writing Vocabulary (a timed word writing task) was used as a dependent variable, giving additional weight to our claim.

Analysis of variance also revealed a statistically significant three-way interaction, which indicated that "closing the gap" could not be fully generalized across all locations where Reading Recovery is implemented. Although seemingly problematic, such an effect was hardly surprising, given the variety of urban, suburban, and rural school districts, with diverse student populations that are characteristic of Reading Recovery sites in the New York metropolitan region. However, when initial differences among students, as expressed in fall scores on the Concepts About Print task, were taken into account, this interaction was no longer significant. (Note: Full Analysis of Covariance results are not reported here, but are available from the authors.)

Such a result, from the analysis of covariance, indicates that individual differences in pre-existing knowledge among students are one possible reason for this site-to-site variation, and not a failed implementation, or an ill fit of Reading Recovery as a literacy intervention in particular sites.

Discussion

This study has reported results from administrations of *An Observation Survey of Early Literacy Achievement* (Clay, 1993a) on 55,231 children. Of these children, 25,601 received Reading Recovery services in first grade during a six-year period from 1992 to 1998. Results have been used to evaluate whether Reading Recovery, as an early literacy intervention, is effective for students who were learning English as another language. In the following sections, we discuss the findings from this study by exploring several issues: (a) English language proficiency as a possible factor in whether children are selected for Reading Recovery services, (b) the relationship between reading achievement and English language proficiency, and (c) limitations and directions for further investigation.

English Language Proficiency as a Factor in Selection

The earlier analysis of success rates for children in Reading Recovery suggests that both native speakers and English language learners are equally likely to be successful and to complete the program. If anything, fluent ESL students are more likely to be successful than native speakers. This is an interesting outcome. Similar findings were summarized by Collier (1989), from studies of children who initially learned two languages simultaneously and outperformed monolingual students in the late elementary years on measures that included linguistic and metalinguistic abilities, cognitive flexibility, and concept formation. In part this is attributed to children's continued cognitive development in both languages.

Reading Recovery tutoring for these fluent ESL students, who initially experienced reading difficulties, may have contributed to language development in English, while other experiences (provided by parents, for example) contributed to continued cognitive development in another language. Such combined cognitive inputs may have allowed these children to begin outperforming their monolingual peers in literacy.

However, our analysis of the scores at the beginning of first grade on *An Observation Survey of Early Literacy Achievement* (Clay, 1993a) suggested that the lowest performing among LEP children were not always selected for Reading Recovery tutoring. In addition, both Fluent ESL, and LEP

children are under-represented in Reading Recovery, with respect to other students "at-risk" in the Comparison Group (see Table 2). As indicated earlier, there may be several practices at the school level shaping such a pattern of results. For example, this pattern may reflect some schools' decisions to delay admission into Reading Recovery for children with English language learners, particularly LEP children, driven by a belief that their English language skills first need to improve to a certain level, before they can be considered for literacy tutoring.

Another practice may be that within a context of limited resources, there is sometimes *pressure* in schools to select students for Reading Recovery for whom progress appears to be more likely, and to exclude those for whom the prognosis appears poor. Anecdotal evidence suggests that both of these practices may reflect a perception among teachers and administrators, that children with limited English proficiency are not suited for Reading Recovery instruction. Whatever the reasons, these practices can lead to decreased opportunities for English language learners to receive the literacy tutoring which would benefit them immediately, according to the data presented in this study. In order to understand these practices more thoroughly, further attention needs to be paid to the effects of other programs and services offered to English language learners, in conjunction with Reading Recovery.

Attempts to predict the reading progress of an individual child initially identified as needing Reading Recovery, suffer from an inherent lack of validity, especially with the low levels of literacy skills that "at-risk" children possess before the first grade. Evidence from Reading Recovery research (Clay, 1993b) demonstrates that it is only after ten weeks in the program that predictions of success can be made with any confidence. Even then predictions still carry a risk of error in at least 30% of cases. Continuous observation and diagnostic teaching (optimally 20 weeks) by a Reading Recovery teacher provides more reliable information on which to make valid and fair assessments on the level of the individual child, particularly when the child's classroom teacher raises questions about the need for referral to special education services.

Since the general pattern of results suggests that Reading Recovery "works" for all students, it is obviously important to ensure that language proficiency does not result in children's inappropriate exclusion from the program. Given the demonstrated effectiveness of the program for all language groups, districts can have confidence that Reading Recovery is an appropriate instructional intervention for these children as well.

Reading Achievement as it Relates to English Language Proficiency

It is evident from the data that Reading Recovery not only contributed to improving the literacy performance of all three language groups (English, Fluent ESL, and LEP), but also reduced the variability in performance among them. Within the Random Sample and the Comparison Groups, however, differences between language groups persisted. At the end of the year, LEP children in both of these groups significantly lagged behind their fluent non-native and native English-speaking peers.

Without an intensive literacy intervention, such as Reading Recovery, non-native English speakers are likely to fall behind by the end of first grade. The data derived from this study indicate that a reading achievement gap exists, both for children initially thought to be "at-risk," and for all other students in first grade. Quality classroom instruction in the primary grades that is tailored to meet diverse learning needs is clearly called for as the first strategy in the prevention of literacy learning difficulties.

However, given the broader research findings on academic achievement in literacy for second language learners (Ramirez et al., 1991), we believe that it is unrealistic to assume that Reading Recovery, as a first grade intervention, can completely protect against the need for further supplementary help. Reading Recovery as an early intervention is designed to reduce the long-term need for remedial reading programs. In the increasingly demanding literacy environment of monolingual English school learning beyond the early grades, school administrators and teachers need to continue to monitor the language and literacy needs of non-native English speakers, and to provide periodic assistance where needed.

Limitations and Directions for Further Research

The general conclusion of this study points to the effectiveness of Reading Recovery tutoring in producing similar outcomes for students with different levels of English proficiency, and offers an appropriate solution for first graders initially experiencing problems in reading and writing. The national Reading Recovery evaluation design, which provided the data for this study, places constraints on the interpretation of the results that are even greater than those typically associated with correlational studies. This is especially true with respect to causality. Issues of program implementation in part determined the selection of students into groups for the purpose of the study. As such, the size of the Comparison Group ("at-risk" students who did not receive the program) was influenced by the level of program implementation in a school. Similarly, some of the clinically valuable measures administered under the Reading Recovery design are ill-suited for statistical analysis due to difficulty level and ceiling effects.

Apparent differences in reading achievement between native English speakers and English language learners may be influenced by other factors, such as characteristics of students (other than native language) and characteristics of their educational environments. Our design takes into account variation across sites, which has typically not been included as a factor in previous studies of Reading Recovery's effectiveness, and eliminates this source of bias from estimates of effects.

Future research studies should take a step further, and try to determine the extent of the influence of specific factors at both student and site levels. Other student characteristics, such as ethnicity, race, socio-economic status, cultural background, and the characteristics of students' native language, are likely to be important factors in the performance of students. Characteristics of Reading Recovery sites as educational environments, such as number of years of Reading Recovery implementation, level of coverage, teachers' experience, urban/suburban location, and district demographics, are also potential explanatory factors for the performance of students at-risk. A convincing case can even be made for the interaction of factors from these two levels (students and sites), especially in a metropolitan area that is characterized by considerable diversity of students.

For example, in some school districts, a number of English language learners may come from populations with relatively high socio-economic status, while native English speakers in some urban districts tend to be of low socio-economic status. In-depth consideration of factors such as these would help evaluation research move beyond general conclusions about the program's effectiveness, and make specific recommendations concerning early literacy intervention for diverse groups of at-risk students. Unfortunately, this diversity is extremely difficult to quantify and control for in this sample of students drawn from sites affiliated with NYU. However large, the sample used in this study lacked adequate

distribution of student characteristics over sites, which is more likely to be found in a national-level sample of Reading Recovery sites and students.

At least some of the issues raised by this study, such as decision-making about which students to admit into Reading Recovery, appear to be related to the characteristics of sites, but it is not possible to explore these hypotheses in great detail from the data at hand. However, a modified national Reading Recovery evaluation design, in place from school year 1998-99, does include additional descriptors on the teacher- and school-level (locale, teacher experience, level of implementation, to mention a few), which will enable more detailed analyses in the future.

Finally, how well the effects of this literacy intervention for English language learners transfer into sustained gains beyond first grade is an issue that remains to be explored. This is crucially important in the light of the fact that English language learners are more likely to be found on the wrong side of the gap in reading achievement, a gap that widens in the course of elementary education and beyond.

Conclusions

Selecting the lowest performing children for Reading Recovery is a key design principle of this program's implementation. We believe that doubts that may exist in some schools about fully adhering to such a principle with respect to English language learners are not supported by the data presented here. The results reported in Table 1 and Figure 1 represent strong evidence that the one-to-one tutoring offered in Reading Recovery constitutes an appropriate setting, in addition to the classroom, to support language and literacy development for children with limited English proficiency.

The substantial database on which we were able to draw allowed us to monitor various aspects of Reading Recovery's implementation and effectiveness. Without such a database across sites, and without the capacity from an external agency to analyze such data (in this case, New York University), identifying potential bias in the delivery of services to English language learners would not have been possible. This speaks in some ways to the value of school-university partnerships in program evaluation.

References

Batelle Institute. (1995). *Longitudinal study of Reading Recovery: 1990-91 through 1993-94*. Ohio Department of Education.

Clay, M. M. (1991). *Becoming literate: The construction of inner control.* Portsmouth, NH: Heinemann.

Clay, M. M. (1993a). *An observation survey of early literacy achievement.* Portsmouth, NH: Heinemann.

Clay, M. M. (1993b). *Reading Recovery: A guidebook for teachers in training.* Portsmouth, NH: Heinemann.

Collier, V. P. (1989). How long? A synthesis of research on academic achievement in a second language. *TESOL Quarterly, 23*, 509-531.

Collier, V. P. (1992). A synthesis of studies examining long-term language minority student data on academic achievement. *Bilingual Research Journal, 16* (1 & 2), 187-212.

Cummins, J. (1986). Empowering minority students: A framework for intervention. *Harvard Education Review, 56*(1), 18-36.

Hobsbaum, A. (1995). Reading Recovery in England. *Literacy Teaching and Learning: An International Journal of Early Reading and Writing, 1*, 21-39.

Hornberger, N. (1992). Biliteracy contexts, continua and contrasts: Policy and curriculum for Cambodian and Puerto Rican students in Philadelphia. *Education and Urban Society, 24*(2), 196-211.

Lucas, T., Henze, R., & Donato, R. (1990). Promoting the success of Latino language-minority students: An exploratory study of six high schools. *Harvard Education Review, 60,* 315-340.

Lyons, C. A. (1998). Reading Recovery in the United States: More than a decade of data. *Literacy Teaching and Learning: An International Journal of Early Reading and Writing, 3*(1), 77-92.

Neal, J. C., & Kelly, P. R. (1999). The success of Reading Recovery for English language learners and Descubriendo La Lectura for bilingual students in California. *Literacy Teaching and Learning: An International Journal of Early Reading and Writing, 4*(2), 81-108.

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. *Reading Research Quarterly, 29*(1), 9-39.

Ramirez, J., Yuen, S., Ramey, D., & Pasta, D. (1991). *Final report: Longitudinal study of structured English immersion strategy, early-exit and late-exit bilingual education programs for language minority children, I & II* (U. S. D. E. Publication No. 300-87-0156). San Mateo, CA: Aguirre International.

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

Snow, C. E., Burns, S. M., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* Washington, D.C.: National Academy Press.

Wilkinson, I. (1998). Dealing with diversity: Achievement gaps in reading literacy among New Zealand students. *Reading Research Quarterly, 33*, 144-167.

Biographical Information

Jane Ashdown is an associate clinical professor in the Department of Teaching and Learning in New York University's School of Education. She coordinates departmental activities with the focus on the continuing professional education of teachers. Dr. Ashdown also directs the Reading Recovery Project and is currently involved in a collaborative research project piloting a cost-effectiveness study of Reading Recovery.

Ognjen Simic is an associate research scientist at New York University's Reading Recovery Project, where he coordinates program evaluation activities in affiliated school districts, and collaborates with project staff on research studies in the field of early literacy intervention. He is working on a doctoral thesis at Fordham University's Psychometric Program in the area of multilevel modeling.