

Examining Oral Reading Fluency Trajectories Among English Language Learners and English Speaking Students

Shane R. Jimerson^{1,*}, Sehee Hong², Scott Stage³, Michael Gerber⁴

¹University of California, Gevirtz Graduate School of Education, California, USA
{jjimerson@education.ucsb.edu}

²Yonsei University, Korea {seeehong@korea.ac}

³North Carolina State University, North Carolina, USA {sastage@ncsu.edu}

⁴University of California, Santa Barbara, California, USA {mgerber@education.ucsb.edu}

Received on 29 November 2012; revised on 29 November 2012; accepted on 5 December 2012; published on 15 January 2013

DOI: 10.7821/naer.2.1.3-11

ABSTRACT

Students' oral reading fluency growth from first through fourth grade was used to predict their achievement on the Stanford Achievement Test (9th ed.; SAT-9 Reading) using a latent growth model. Two conditional variables related to student status were used to determine the effects on reading performance - English language learners (ELLs) with low socioeconomic status and low socioeconomic (SES) status alone. Results revealed that both types of student status variables reliably predicted low performance on initial first grade oral reading fluency, which later predicted fourth grade performance on the SAT-9. However, the reading fluency trajectories of the ELLs and monolingual English students were not significantly different. In addition, when both student status variables and letter naming fluency were used to predict initial oral reading fluency, letter naming fluency dominated the prediction equation, suggesting that an initial pre-reading skill, letter naming fluency, better explained fourth grade performance on the SAT-9 than either ELL with low SES or low SES alone. The discussion focuses on how to better enable these readers and how oral reading fluency progress monitoring can be used to assist school personnel in determining which students need additional instructional assistance.

KEYWORDS: ORAL ENGLISH, ORAL READING FLUENCY, SECOND LANGUAGE LEARNING, FOREIGN LANGUAGE LEARNING, LANGUAGE FLUENCY

Children from Spanish-speaking countries (i.e. Hispanics¹) comprise the largest minority group among youth (U.S. Department of Education, 2003). For instance, in California, approximately one-third of the residents are Hispanic (Cauce & Domenech-Rodriguez, 2002). It has been estimated that the majority of the language minority population in the United States is Spanish-speaking (Greenberg, Macias, Rhodes, & Chan, 2001). The recent expansion of Hispanic population in the United States is reflected in the 100% increase of Spanish-speaking students receiving bilingual and ESL services since 1984 (Bentz & Pavri, 2001).

As a group, Spanish-speaking English language learning students are at an increased risk of having academic problems. The confluence of risk factors including low socioeconomic status and, more importantly, limited access to bilingual education programs due to resource limitations or policy (e.g. Propositions 227 in California, 203 in Arizona; both mandate English only instruction), have been associated with lower scores in reading, mathematics, and science for Hispanic students relative to their Anglo peers and increased school dropout rates (U.S. Department of Education, 2003). Given the aforementioned factors, Hispanic students who are also English Language Learners (ELL) but come to be placed in English-only programs perceivably find themselves at an increased risk level for learning problems, especially as regards learning to read in English.

The Curriculum-Based Measurement (CBM) methodology as used to assess reading has proved to be a valuable strategy for assessing progress in skilled reading by students (Deno, 1985; Deno & Fuchs, 1987; Fuchs & Deno, 1991; Powell-Smith & Bradley-Klug, 2001; Shinn, 1989). The CBM methodology, which generally consists of students reading aloud for one minute from an appropriate passage yielding an oral fluency score (i.e. number of words read correctly per minute), has two highly relevant uses. Firstly, individual reading probes have been measured at a single point in time for purposes of screening and identifying relative rank in comparison to the group as in normative testing (for instance, as described by Habendank-Stewart & Kaminski, 2002; Shinn, 1989); and secondly, repeated measures have been used to measure growth over time in oral reading fluency (for instance, as described by Shinn, 1989). CBM stands out as one of the few brief forms of measurement with a large body of data supporting its technical adequacy and practical application to children assessment in elementary grades (Howell, Kurns, & Antil, 2002; Fuchs & Deno, 1991, 1994; Marston & Magnusson, 1985; Shinn, 1989; Tindal, 1993).

*To whom correspondence should be addressed:
University of California
Gevirtz Graduate School of Education
Counseling, Clinical, and School Psychology
93106-9490 Santa Barbara
California, USA

¹The definition of the term "Hispanic", as used herein, is consistent with that used by the U.S. Census Bureau.

1 CBM AMONG ENGLISH LANGUAGE LEARNERS

Although many facets of CBM have been investigated with monolingual English speaking students, its use with ELLs has received relatively scant empirical attention (Baker & Good, 1995; Baker, Plasencia-Pienado, & Lytle, 1998; Bentz & Pavri, 2001). The extant literature does provide preliminary evidence that CBM can be used to examine reading skills of language-minority students (Baker & Good, 1995). However, no previous research has been conducted to investigate growth specifically over the primary school grades of English Language Learners' oral reading fluency in English and its relationship to high stakes reading tests.

In addressing the perception of increased academic risk among ELLs, it is important to determine expected parameters of literacy growth under typical instruction conditions (Baker *et al.* 1998). While estimates of individual and normative academic progress indices have long been touted as a CBM methodology strength (Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993; Hasbrouck & Tindal, 1992; Shinn, 1989), further research is necessary to ascertain its utility and validity to estimate similar indices for ELLs. For instance, in a seminal study about students with reading disabilities, Francis, Shaywitz, Stuebing, Shaywitz and Fletcher (1996) investigated whether the developmental lag versus the deficit model better explained the growth of students with reading disabilities compared to typically developing readers. A similar hypothesis test might apply to ELLs' reading development. Specifically, the developmental lag hypothesis suggests that learners with reading difficulties lag behind their peers due to delayed development of those reading sub-skills which are critical for reading acquisition. If this hypothesis were true, one would expect ELLs' growth trajectories to differ from that of monolingual, typically developing readers - the underlying assumption being that, with the eventual development of necessary sub-skills, ELLs will catch up with typical monolingual readers. The deficit model conversely posits that initial deficits remain throughout students' schooling, and the only way to alter the course of students' achievement trajectories is through an intervention targeted at the non-existing sub-skills.

The current study examines the developmental lag versus deficit hypothesis as it may relate to ELL beginning readers. In addition, four explicit questions are examined seeking to clarify the nature of ELL perceived risk: (a) Does a student's relative standing in first grade oral reading fluency predict fourth grade performance on a high stakes test (i.e. SAT-9, Total Reading)? (b) Does growth from first grade to fourth grade in oral reading fluency predict 4th grade reading achievement on SAT-9? (c) Does the combination of ELL and low SES and/or low SES status alone have a direct effect on SAT-9? (d) Does the combination of ELL and low SES and/or low SES alone have an indirect effect on SAT-9 through first grade reading level and/or growth across first through fourth grade on oral reading fluency?

2 METHOD

2.1 Participants

Longitudinal data collected from students, beginning in first grade and continuing through fourth grade, in a Southern California school district served to examine developmental trajectories in oral reading fluency. Students included both low

SES ELLs (i.e. those who received free or reduced lunch; $n = 85$) and similarly defined low SES monolingual English speaking students (i.e. $n = 70$). Each year, all students took part in a continuous progress monitoring reading fluency assessment during the first and last month of the academic year.

3 MEASURES

Oral Reading Assessment Level by Jimerson (ORAL-J). The ORAL-J (Jimerson, 1997, 2000) was designed with a view to expand the use of oral reading probes to all students school-wide or district-wide. The ORAL-J includes an oral fluency sub-test where scores on three reading passages are averaged to yield the words per minute (WPM) score. The methodology is consistent with previous research on the curriculum-based measurement of reading (Good & Jefferson, 1998; Hintze, Callahan III, Matthews, Williams, & Tobin, 2002; Shinn, 1998). The technical data for the ORAL-J (Jimerson, 2000) demonstrate both strong reliability and validity as a general reading ability measure. ORAL-J reliability has been demonstrated in several ways with various grades over the past few years. Reliability analyses have been conducted in two formats: test-retest and initial trainings. Test-retest data (that is, 68 first, second, and third grade students in the same school district) indicate a high correlation for word fluency (that is, $r = .96$). In our view, this strong correlation suggests that the ORAL-J is a reliable measure across different assessors at different times for this population. The data comparing assessors' scores during the initial training session with the predetermined consensus scores also reveal a strong reliability ($r = .94$). In addition to word fluency, letter naming fluency was equally used in one of the analyses carried out in the current study. Analyses of letter naming fluency have yielded high reliability correlations similar to word fluency ($r = .93$).

Stanford Achievement Test – Ninth Edition (SAT-9). During the spring of each year, the Stanford Achievement Test – Ninth Edition, or SAT-9 (Harcourt Brace & Company, 1997a) was administered as well. The Total Reading scale utilized in this research includes the Reading Vocabulary sub-test and the Reading Comprehension sub-test. Both sub-tests were included in this study because CBM related oral reading probes are commonly seen to predict general reading proficiency measures. SAT-9 was normed on approximately 250,000 students, sampled in a manner that permitted to be consistent with the national school population in terms of region, ethnicity, socioeconomic status, and urbanicity (Harcourt Brace & Company, 1997a). There is extensive research work reporting SAT-9 reliability and validity (Harcourt Brace & Company, 1997b).

Student Information. The demographic variables of interest in the present analyses are ethnicity, socioeconomic status, and English language proficiency. Ethnicity and socioeconomic status variables were determined on the basis of information provided by parents to school officials upon enrolling their children in school. The ethnic groups included in this study represented both Hispanic and Anglo students. Students from other ethnic groups were excluded from all analyses because the samples were too small to be statistically meaningful. Socioeconomic status was dichotomized using the district's data on students who were eligible for free or reduced-cost lunch, according to federal standards. English language proficiency classification was based on the district administered Language Assessment Scale (LAS) (De Avila & Duncan, 1984), with students being classified as an English Language Learner if they obtained an overall score of 1-3 (non-speaker – limited speaker

and – emergent speaker) on the test. During the study period, the district used the McGraw-Hill Open Court reading curriculum and instruction was conducted in English following California proposition 227.

3.1 Data Collection Procedures

The ORAL-J was administered to students on a one-to-one basis by classroom teachers, each one of whom had been previously trained and certified by the first author. Testing of all students was completed within a period of two weeks, both at the beginning and at the end of the school year. SAT-9 was administered in a group setting during a two-week period in the spring of each year beginning in second grade. The results of test administrations between spring 1998 and spring 2001 provide the data used in this study.

4 ANALYSES

Latent Growth Modeling (LGM) statistical procedures were utilized to examine and compare reading growth trajectories among ELLs and monolingual English speaking students. LGM includes appropriate statistical procedures for examining quantitative change or heterogeneity in growth where there is expected change in an outcome variable, and not only describes each individual's growth but also permits the analysis of individual differences in these trajectories over time (McArdle & Epstein, 1987; Meredith & Tisak, 1990). If, for example, these trajectories produced a collection of straight lines for a group, LGM would reflect individual differences in the slopes and intercepts of those lines (Duncan & Duncan, 1994). In the linear change model, LGM can be expressed as:

$$y_{it} = \eta_{0i} + \eta_{1i}x_t + e_{it}$$

$$\eta_{0i} = \beta_0 + u_{0i}, \eta_{1i} = \beta_1 + u_{1i}$$

where y_{it} represents the measure of response variable y (i.e. oral reading fluency) for individual i at time t and x_t represents the measure of time t (i.e. spring of each academic year). The intercept for individual i is η_{0i} where i represents reading fluency (i.e. number of words read correctly at the end of first grade), η_{1i} is the slope for individual i (i.e. growth in the number of words read correctly from the end of first grade to the end of fourth grade), and e_{it} is the residual for individual i at time t . β_0 is the average intercept value (i.e. the first grade average number of words read for all 247 students), β_1 is the average slope value, and u is the residual.

The three equations above are typical for Structural Equation Modeling (SEM), the first one representing a measurement model while the other two represent a structural model (particularly when correlates of change are incorporated to account for individual differences in change). In LGM, the measurement model represents individual change over time while the structural model represents inter-individual differences in change.

In the measurement model, y_{it} is predicted by η_{0i} and η_{1i} , which are the underlying factors standing for initial status (intercept factor) and rate of change (slope factor), respectively. The mean of the intercept factor represents the group initial status parameter and its variance reflects individual differences in initial status. The mean of the slope factor represents the

group growth parameter and its variance reflects individual differences in change over time.

The following procedures were used to conduct data analyses in the current study. First, a well-fitting LGM was estimated, where growth of reading fluency is modeled using the repeated words per minute correct as the metric. Because this LGM does not include any predictors, this model is called 'unconditional.' Then, a conditional LGM was developed through the incorporation of the other variables that influence or are influenced by the intercept and slope factors. Figure 1 shows the conditional LGM, through which key issues in this study were addressed.

Since listwise/pairwise deletion can result in biased parameter estimates due to non-random attrition (Arbuckle, 1996), this study employed full information maximum likelihood (FIML) estimates. This type of estimate has proved to be very efficient for incomplete data under the assumption that students' unavailability at some testing times was random (Schafer & Olsen, 1998).

Model evaluation criteria. Because the chi-square fit index is highly sensitive to sample size, a decision was made to use the non-normed fit index (NNFI; Bentler & Bonett, 1980) and the root mean square error of approximation (RMSEA; Steiger & Lind, 1980), both of which are relatively independent of sample size. Another reason for choosing the indexes was that both of them take into account model complexity, which is an important property when comparing several alternative models with different degrees of complexity. Values of .90 and .95 can be defined as acceptable and good fit, respectively, to assess NNFI fit (Hu & Bentler, 1999). Values of RMSEA of .05 indicate close fit, values in the vicinity of .08 indicate fair fit, and values of .10 and larger indicate poor fit (Browne & Cudeck, 1993).

Unconditional LGM. As noted above, an unconditional LGM was firstly developed to identify the growth functional form of correct words per minute correct. The intercept factor loadings in this LGM were set to 1, which represents the starting point of the growth curve at Time 1 with the slope factor loadings set to 0, 1, 2, and 3 for four time points to represent the linear growth function.

A non-linear model was equally developed where the slope factor loadings were set to 0, *, *, and 1 for four time points (the symbol * indicates a freely estimated parameter). The last factor loading in this model was fixed at 1 for identification purposes but the second and third coefficients were estimated from the data. The comparison between the linear and non-linear models allowed us to check whether the growth pattern was linear or nonlinear (see Table 1). Since the linear model is nested within the nonlinear one, a χ^2 difference test was performed. According to the test, the χ^2 difference (i.e. 16.36-5.66 = 10.70) was statistically significant with the difference in degrees of freedom (i.e. 5-3 = 2), which indicates that the non-linear model fits the data significantly better than the linear one. The nonlinear model shows a better fit too in terms of NNFI and RMSEA, thus suggesting that the growth pattern is not linear.

Table 1. Model Fit of Latent Growth Models

Model	χ^2	df	NNFI	RMSEA (90% Confidence Interval)
Linear Model	16.36	5	0.89	0.08 (0.04, 0.12)
Non-linear Model	5.66	3	0.91	0.05 (0.01, 0.09)
Conditional Model	11.88	9	0.94	(0.00, 0.06)

Hierarchical Multiple Regression. A hierarchical multiple regression analysis was conducted as a follow-up analysis to the LGM. This regression analysis used two blocks of variables to predict oral reading fluency in the spring of first-grade. The first block of variables was ELL with low SES versus low SES alone. The second block included both ELL with low SES vs. low SES and first grade letter naming fluency.

5 RESULTS

Means and standard deviations for both groups are summarized in Table 2. The parameter estimates obtained from the nonlinear LGM are offered in Table 3. The mean of the intercept factor presented in Table 3 is 49.27, which represents the initial status. The mean of the slope factor is 80.04, which indicates the amount of growth during the periods of spring 1998 through spring 2001. The loadings for the slope factor show the amount of growth during each interval. That is, the words per minute correct were increased by 45.8% of 80.04, or 36.66, between spring 1998 and spring 1999 and by 74.9% of 80.04, or 59.39, between spring 1998 and spring 2000. Variances of the intercept and slope factors were statistically significant, which indicates that significant individual differences existed in both initial status and change over time. The correlation between the intercept and slope factors was also significant ($r = -0.213$). The negative correlation means that individuals who had lower scores at the beginning showed higher growth rates over time. Figure 1 demonstrates two growth shapes: one developed using the actual data and the other developed using the parameter estimates from the nonlinear LGM. The graph illustrates that the model fits the data quite well.

Conditional LGM. After identifying the fit for the unconditional LGM, a conditional LGM was developed to test key issues in this study. Figure 2 illustrates the conditional model tested. As shown in figure 2, the fit of the conditional model in terms of NNFI (= .94) and RMSEA (= .03) was acceptable. The parameter estimates obtained from the conditional LGM are also presented in the figure.

Does the student's relative standing in first grade oral reading fluency (i.e. ORAL-J, 2000) predict fourth grade performance on a high stakes test (i.e. SAT-9, Total Reading)? The standardized estimate of first grade oral reading fluency effects on the fourth grade standardized overall reading achievement was .663, indicating a large effect. Therefore, first-grade oral reading fluency strongly predicted reading performance in fourth grade on the SAT-9 high stakes test.

Does growth from first grade to fourth grade in oral reading fluency predict performance on SAT-9? The standardized estimate of growth in oral reading fluency from first to fourth grade in relationship to reading performance on a high stakes test was .254, indicating that growth in oral reading fluency had a moderate effect, although not as strong as initial first grade oral reading performance.

Table 2. English Language Learners and English speaking students on Oral Reading Fluency and SAT-9.

	Monolingual English Students		English Language Learners	
	M	SD	M	SD
Oral Reading Fluency 1st Grade	64.46	31.79	40.65	28.83
Oral Reading Fluency 2nd Grade	101.59	36.81	74.93	33.04
Oral Reading Fluency 3rd Grade	90.24	35.58	63.93	29.88
Oral Reading Fluency 4th Grade	146.87	32.02	119.80	28.52
Oral Reading Fluency Growth Curve Slope Grades (1-4)	12.80	5.64	12.79	4.45
Total Reading 4th Grade (SAT-9)	68.54	17.42	46.04	18.19
Word / Vocabulary 4th Grade (SAT-9)	70.25	16.79	46.13	20.14
Sentence Completion 4th Grade (SAT-9)	66.06	17.41	46.77	16.65

Note: Based on Data from grades 1st-4th.

Table 3. Parameter Estimates of the Non-linear LGM

Regression Weights	Estimate
W1S98 <----- INT	1000
W2S99 <----- INT	1000
W3S00 <----- INT	1000
W4S01 <----- INT	1000
W1S98 <----- SLP	0.000
W2S99 <----- SLP	0.458*
W3S00 <----- SLP	0.739*
W4S01 <----- SLP	1000
Means	INT 49.276*
	SLP 80.047*
Correlations	INT <--> SLP -0.239*
Variances	INT 1053.361*
	SLP 289.153*

Note: * indicates "significant" at $\alpha = .05$.

W1S98 = First Grade Oral Reading Fluency (spring)

W2S99 = Second Grade Oral Reading Fluency (spring)

W3S00 = Third Grade Oral Reading Fluency (spring)

W4S01 = Fourth Grade Oral Reading Fluency (spring)

INT = First Grade Oral Reading Fluency (spring)

SLP = First Grade Oral Reading Fluency (spring) through Fourth Grade Oral Reading Fluency (spring)

Figure 1. Illustration of the Model Fit with the Actual Data

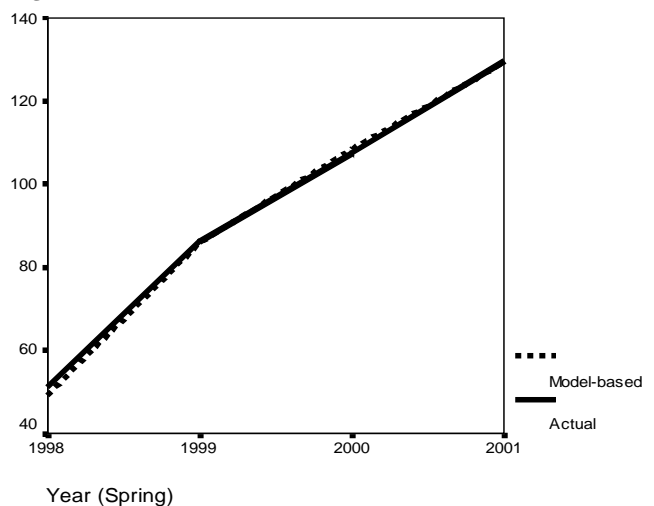
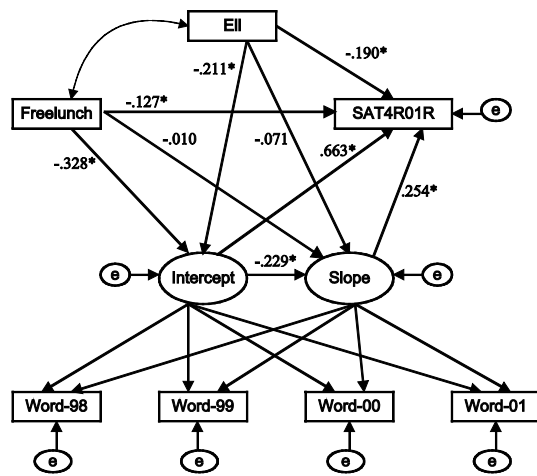


Figure 2. Conditional LGM with Parameter Estimates: Representing the Prediction of SAT-9 Scores in 4th Grade from Oral Reading Fluency Scores Across Grades 1st - 4th, Considering Free and Reduced Lunch among English Language Learner Students.



Note:

ELL = English Language Learner Status 0/1
 Free Lunch = Free Lunch Status 0/1
 SAT4R01R = SAT-9 Fourth grade Reading Score (spring)
 Word-98 = First Grade Oral Reading Fluency (spring)
 Word-99 = Second Grade Oral Reading Fluency (spring)
 Word-00 = Third Grade Oral Reading Fluency (spring)
 Word-01 = Fourth Grade Oral Reading Fluency (spring)
 Numbers in parentheses = Parameter Standardized Estimates
 * indicates "significant" at $\alpha = .05$
 NNFI = .94 & RMSEA = .03

Does the combination of ELL and low socio-economic status (SES) and/or low SES status alone have a direct effect on SAT-9? The combined status of ELL and low SES had a low direct effect ($-.127$) on SAT-9, and low SES status alone had a low direct effect ($-.190$) on SAT-9 too. These two results taken together indicate that low SES negatively impacted fourth grade performance on a high stakes test, but also that the addition of ELL status did not further decrement students' reading performance.

Does the combination of ELL and low SES and/or low SES alone have an indirect effect on SAT-9 through first grade reading level and/or growth across first through fourth grade on oral reading fluency? The indirect effect of the combined status of ELL and low SES had a large effect on fourth grade reading performance through its indirect effect on initial first grade oral reading fluency ($-.211$) to SAT-9 (.663) performance in fourth grade. Likewise, monolingual low-SES status demonstrated a similarly large indirect effect on fourth grade reading performance through its indirect effect on initial first grade oral reading fluency ($-.328$) to SAT-9 (.663) performance in fourth grade. The relative difference between the combined status of ELL and low SES versus low SES alone was not statistically significant in relationship to initial first grade oral reading fluency ($t[153] = .089, p = .92$). However, there was no indirect effect through slope as neither ELL plus low SES nor low SES status alone was conditionally related to slope in oral reading status.

Overall model results. Overall, the results indicate that the initial first grade oral reading ability was the best estimate of fourth grade reading performance on a high stakes reading test. In fact, the conditional estimate of initial reading achievement to

SAT-9 was of larger magnitude than the absolute sum of all the other conditional estimates within the model (i.e. ELL and low SES = $-.127$, low SES = $-.190$, oral reading fluency slope = $.254$ versus initial oral reading fluency = $.663$). Growth in oral reading fluency across the first through fourth grades also moderately estimated fourth grade reading performance. The results also indicated an indirect effect of low SES status negatively affecting initial first grade oral reading fluency ($-.328$) which in turn strongly predicted fourth grade reading achievement measured by SAT-9 (.663). These results compare similarly to the results of conditional estimates for the combined status of ELL and low SES.

Hierarchical Multiple Regression. The first block of variables assessed in the hierarchical multiple regression analysis was the combination of ELL plus low SES and low SES alone. With the data from 308 students used in this analysis, both variables were significantly related to oral reading fluency (i.e. $F = 23.522, p < .0001, R^2 = .134$; ELL + low SES, $\chi^2 = -.352, t = -6.4, p < .0001$; low SES, $\chi^2 = -.218, t = -3.96, p < .0001$). When both variables were once again entered with letter-naming fluency, which was measured in the fall of first-grade, the results showed that only low SES and letter-naming fluency remained statistically significant in their relationship to oral reading fluency (i.e. $F = 115.1, p < .0001, R^2 = .532$; ELL x low SES, $\chi^2 = -.026, t = -.574, p = .567$; low SES, $\chi^2 = -.091, t = -2.20, p = .028$; letter naming fluency, $\chi^2 = .707, t = 16.079, p < .0001$). Regression results indicate that the best prediction of initial oral reading fluency measured at the end of first grade was letter naming fluency when compared to both ELL and low SES combined and low SES as a single measure.

Additional descriptive data (Table 4) reveal similar correlations found between ELLs and monolingual English speaking students for reading fluency scores in relation to current and projected reading proficiency. For example, the oral reading fluency scores from the spring of 2001 for both ELLs and monolingual English speaking students in the fourth grade strongly correlate with the SAT-9 scores of these same students in the fourth grade (2001). These correlations, .72 and .62, respectively, demonstrate the similar associations between reading fluency scores and current reading proficiency across these samples of students. Furthermore, oral reading fluency scores for both ELLs and monolingual English speaking students in the first grade (spring of 1998) turned out to have similarly strong correlations with the fourth grade scores on SAT-9 obtained by the same students four years later, in 2001: .66 and .53, respectively.

Table 4. Correlations between Reading Fluency and SAT-9 4th Grade Total Reading

	Monolingual English Students	English Language Learners
First Grade Fall 98	.46	.41
First Grade Spring 98	.53	.66
Second Grade Fall 98	.55	.74
Second Grade Spring 99	.57	.78
Third Grade Fall 99	.61	.68
Third Grade Spring 00	.68	.76
Fourth Grade Fall 00	.64	.71
Fourth Grade Spring 01	.62	.72

All correlations significant at $p < .01$

6 DISCUSSION

The primary purpose of this study was to examine reading trajectories among English Language Learners (ELLs) and monolingual English speaking students considering the developmental lag versus deficit models (Francis *et al.*, 1996) as a frame of reference. The conditional impact resulting from the combination of ELL and low SES versus low SES status alone on initial first grade oral reading fluency and growth in oral reading fluency across first through fourth grade was tested for their relationship with a high stakes reading test, the Stanford Achievement Test-Total Reading (9th ed.).

The conditional model results support the deficit model in that relative standing in first grade oral reading fluency was predictive of their standing in fourth grade reading achievement for both groups of students, whereas their growth in oral reading fluency was not. The standardized estimate from low SES to initial first grade oral reading fluency was $-.328$, which suggests that about a third of these students on average started below the median. However, it is important to note that students with the combined status of ELL and low SES were no more likely than their monolingual counterparts to be any further behind, which suggests that, given a low SES, the addition of ELL status did not further negatively impact overall standing. The results indicate that the students did not fall further behind as the Mathews Effect would have suggested (Stanovich, 1986). The Mathews Effect is paraphrased as the "the rich get richer and the poor get poorer." Evidence of this would have been that the more capable readers would have shown greater growth over time than poorer readers, who would lagged further behind in growth over time with a statistically significant relationship between ELL plus low SES and growth in oral reading fluency. Interestingly, this same finding indicates that the "gap" between lower and higher performing was not closed. Further discussion is presented for each of the four questions, along with an analysis of additional information that better frames assessments examining sub-skills for students who have difficulty learning to read.

Does the student's relative standing in first grade oral reading fluency (i.e. ORAL-J, 2000) predict fourth grade performance on a high stakes test (i.e. SAT-9, Total Reading)? First-grade oral reading fluency strongly predicted reading performance in fourth grade on a high stakes standardized test. Another study which used a growth curve analysis with fourth-graders' oral reading fluency and performance on a high stakes reading test in fourth grade also showed a significant relationship between initial oral reading fluency and a high stakes reading test (Stage & Jacobsen, 2001). Jointly considered, these studies indicate that a single oral reading fluency measurement is predictive of later reading performance on high stakes tests of reading performance. Therefore, students performing in the lower percentile ranks in initial oral reading fluency are likely to show poor outcome on high stakes tests without specific intervention.

Does growth from first grade to fourth grade in oral reading fluency predict performance on SAT-9? The standardized estimate of growth in oral reading fluency from first to fourth grade had a moderate effect on high stakes test performance (.254), although the effect was half the effect that initial oral reading fluency was (.663). Two important considerations need to be made with regard to the first finding. Firstly, the average growth over time was decelerating slightly, suggesting that the growth in oral reading fluency was not linear. In particular, their

oral reading fluency growth started to asymptote as students approached third and fourth grade. Other research has shown this trend as well (e.g. Fuchs *et al.*, 1993). The development of reading skills suggests that as students gain word reading fluency in the middle primary grades, oral reading fluency becomes less critical once a certain level of fluency is achieved and then reading comprehension becomes the most salient reading skill (Adams, 1990). However, in one of the original demonstrations where oral reading fluency was used for progress monitoring, Marston (1987) noted the importance of using the CBM technology for students with Special Education eligibility because growth could be evidenced. In another example, Stage (2001) monitored a sample of low SES second grade-students whose oral reading fluency placed them in the lowest quartile across the year and found that the students continued to remain there throughout the year. At the end of the school year, these students were enrolled in an intensive six-week summer school reading program. By the end of the program, 71% of the students in the summer program were reading at a level within second to third quartile of the typical student. The results of the current study suggest the same trend in the data; in other words, once the student was identified as performing in the lower third of the distribution, they were more likely to stay at this relative rank and would require a more intense intervention in order to change their trajectory.

Does the combination of ELL and low socio-economic status (SES) and/or low SES status alone have a direct effect on SAT-9? As noted above, the combined status of ELL and low SES had a limited direct effect on fourth grade reading performance and so did low SES status alone. When statistically compared to the indirect effects through oral reading fluency, the direct effects of ELL combined with low SES and low SES alone accounted for half of the influence exerted by the indirect effects, which suggests a relatively weak impact.

Does the combination of ELL and low SES and/or low SES alone have an indirect effect on SAT-9 through first grade reading level and/or growth across first through fourth grade on oral reading fluency? The indirect effect of the combined status of ELL and low SES had a large effect on fourth grade reading performance through its indirect effect on initial first grade oral reading fluency to reading performance in fourth grade and likewise for the monolingual low SES status. As mentioned in the overall model findings, children raised in low economic status environments are more likely to be delayed in language development and also less likely to have access to preschool programs that explicitly teach pre-reading skills (Whitehurst & Fischel, 2000). This further suggests that screening for pre-reading skills can further identify students who are apt to have difficulty learning to read. For example, kindergarten letter naming fluency significantly predicted first grade oral reading fluency in a sample of ethnically diverse students (Stage, Sheppard, Davidson, & Browning, 2001) and was found to be the best concurrent Dynamic Indicator of Basic Early Literacy Skills measure of word reading ability in kindergarten children (Elliott, Lee, & Tollefson, 2001). In the current study, first-grade fall letter naming fluency represented the best predictor of first grade spring oral reading fluency in comparison to the combined status of ELL and low SES or low SES status alone. As a matter of fact, it explained 40% of variance in spring oral reading fluency. These findings have important implications for practice, as they indicate that early pre-reading skills are better markers of students' reading development than are ELL or low SES status.

7 LIMITATIONS AND FUTURE DIRECTIONS

The results obtained in the current study support how useful the CBM methodology can be with low SES students in grades one through four regardless of their initial English language proficiency. The results should be interpreted considering a number of limitations, though. Firstly, with respect to developmental trajectories, as it was found that both the LGM and correlational analyses actually reveal a similar trend among both populations when examining current and future reading performance. Although no differences appeared in reading growth trajectories between the two groups, it is premature to declare that developmental reading trajectories will be the same for both groups. The findings reported here should be considered preliminary in terms of examining second language reading. The students in this study were enrolled in an English only school district that does not assess students in Spanish annually. The school district is only mandated by law to assess students in their primary language when they are initially identified as English Language Learners. The district does not currently have an obligation to assess ELLs annually in their primary language looking at their Spanish language development. It is difficult to surmise their progress in a second language when data on their primary language development are not available.

Secondly, because the authors used a generic language classification (English Language Learner), it is more difficult to examine a second language reading trajectory. While normative indices of academic progress for monolingual English speaking students have long been established (Fuchs *et al.*, 1993; Hasbrouck & Tindal, 1992; Shinn, 1989) additional evidence is needed for a critical examination of these same indices when acquiring a second language and learning to read. Future research should be conducted that utilizes English language classification in greater detail (i.e. beginning; early intermediate; and intermediate proficiency levels). This information could provide more insight into whether the same developmental trend found in this study holds constant when adding different English language proficiency levels as determined by standardized tests. Normative indices for populations that are learning to read in a second language will start to be supplied as well.

Thirdly, given the current sociopolitical zeitgeist, namely No Child Left Behind and Propositions 227 and 203 (advocating for English only instruction), it is particularly important to have an appropriate methodology that can be used with all students regardless of their categorical differences. Despite the evidence from this study that CBM methodology can produce reliably predictive measures of Hispanic student's later English reading achievement, concerns still exist. For example, decades of research have unequivocally provided evidence of the relationship between monolingual English speakers' reading fluency and reading proficiency. However, this relationship is less clear when examining Spanish speaking among English Language Learners. The unique process of acquiring a second language will look different in terms of rates of English language acquisition. Thus, more research is needed that utilizes this same methodology, including assessments that can be administered in two languages to better understand primary as well as second language reading development.

Lastly, different reading passages were used over time in this study. Hintze *et al.* (2002) argued that utilizing passages based on grade level appropriateness rather than using the same ones over time increases the likelihood of other variables having a greater chance of explaining significant amounts of variance.

While the current authors and others (Klein & Jimerson, in press; Kranzler, Miller, & Jordan, 1999) have used grade level appropriate passages and found no evidence of bias, the current findings actually provide evidence in favor of grade level appropriate passages. Additional research is therefore needed to clarify this further.

The results of this study reveal that oral reading fluency as measured in a CBM based progress monitoring strategy is a methodology that can be used with diverse populations, namely Hispanic, Spanish speaking students. This study should be replicated with similar and diverse populations, particularly in the context of transitional or dual language immersion programs, in order to better understand the developmental trajectories of English Language Learners learning to read in a second language.

8 SUMMARY

Acknowledging the current emphasis on accountability within the educational system, continuous progress monitoring of oral reading fluency may be reliably used to identify current reading skills, which are highly associated with future reading performance and provide an opportunity to evaluate students' reading progress over time. The results of this study reveal that the early oral reading fluency assessment used for this longitudinal project does reveal similar developmental reading acquisition patterns and predictions across ELLs and monolingual English students. English Language Learner students demonstrated similar growth patterns to those of monolingual English speakers in reading development across time. Further research is necessary to examine the developmental reading progression of ELLs based on collecting data in both languages over time, from dual language immersion or transitional bilingual programs.

9 CONCLUSION

Given that the student population in the United States is becoming increasingly diverse, it becomes particularly important to have an unbiased measure that can be used to: 1) monitor student reading progress; and 2) utilize a metric that can provide information about future reading performance for all students. To date, little empirical research has demonstrated CBM as an unbiased form of assessment with other ethnic groups. Hintze *et al.* (2002) found that CBM appears to be a sensitive form of reading assessment for African Americans. The current study offers additional empirical research according to which language classification was not associated with reading growth trajectories. Thus, brief fluency probes can provide valuable information about reading with diverse populations.

Baker and Good (1995) provided preliminary evidence supporting the validity of CBM English reading measures as a measure of English reading proficiency and reading comprehension for bilingual Hispanic students. The findings reported here expand upon their pioneering work. In addition, the unique contribution of the present study lies in its relatively large sample size and the longitudinal data collected over a four-year period. The few studies that exist typically utilize small sample sizes and data are generated over a short time interval. Moreover, there are few studies to date that have employed rigorous statistical analyses such as LGM modeling. This statistical procedure is the most appropriate one to use when examining growth trajectories and individual variation.

Developmental trajectories cannot be fully understood without employing this type of methodology.

In comparing and contrasting the association of reading fluency scores and current reading proficiency, similar associations between reading fluency scores and current reading proficiency were found for both populations. This supports the concurrent validity of oral fluency across ELLs and monolingual English students. Furthermore, no significant differences appeared between the groups in the association of reading fluency scores and future reading proficiency, which indicates a consistent association of reading fluency scores among grades first through fourth for both monolingual English and ELLs. Students' scores on the oral reading fluency assessment in the early grades were associated with later performance on the fourth grade SAT-9 test. Data revealed positive associations indicative of high predictive validity. Specifically, those students (both monolingual English and ELLs) who achieved a higher reading proficiency level in first grade performed better on SAT-9 in fourth grade. These findings emphasize the potential utility of this early reading fluency assessment with diverse populations to accomplish the important task of facilitating the educational success of all students. This study provides current information about the reading acquisition patterns and developmental trajectories of both monolingual English and ELLs. In this period characterized by the emphasis on educational standards and accountability, the early identification and progress monitoring of reading skills among all students warrants further consideration by educational professionals. This study emphasizes that letter fluency, sound fluency, and oral reading fluency are important and practical indicators of both current and future reading success for all students.

Furthermore, the data stemming from the current study are also consistent with the extant literature, which suggests that the metric of words per minute is valuable in terms of predicting future reading performance (Hintze *et al.*, 2002; Sibley, Biwer, & Hesch, 2001). Although this methodology has often been used to monitor progress in reading development, there is scant research that addresses its utility to predict future reading performance. Sibley *et al.* (2001) examined oral reading fluency measures to predict student performance on high stakes assessments (such as the Illinois Standards Achievement Test (ISAT)) and Stage and Jacobsen (2001) paid attention to the relationship between oral reading fluency and the Washington Assessment of Student Learning. Results support the utility of reading fluency benchmarks for predicting student performance on subsequent high stakes tests. The present study found that the words per minute metric predicted future reading performance on a high stakes test, SAT-9. The data presented herein demonstrated similar predictive validity of oral reading fluency scores among ELLs and monolingual English speaking students. In particular, initial scores, reading trajectories and language classification (English learner) predicted SAT-9 scores. Collectively, these findings provide strong evidence supporting the relative utility of this methodology with diverse populations.

REFERENCES

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: The MIT Press.
- Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G.A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Baker, S. K., & Good, R. (1995). Curriculum-based measurement of English reading with bilingual Hispanic students: A validation study with second-grade students. *School Psychology Review*, 24(4), 561-578.
- Baker, S. K., Plasencia-Peinado, J., & Lezcano-Lytle, V. (1998). The use of curriculum-based measurement with language-minority students. In M.R. Shinn (Ed.), *Advanced Applications of Curriculum-Based Measurement* (pp. 175-213). New York: The Guilford Press.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606. doi:10.1037/0033-2909.88.3.588
- Bentz, J., & Pavri, S. (2001). Curriculum-based measurement in assessing bilingual students: A promising new direction. *Diagnostique*, 25(3), 229-248.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- Cauce, A. M., & Domenech-Rodriguez, M. (2002). Latino families: Myths and realities. In J. M. Contreras, K. A. Kerns & A. M. Neal-Barnett (Eds.), *Latino Children and Families in the United States: Current Research and Future Directions* (pp. 3-25). Westport, CT: Praeger.
- De Avila, E. A., & Duncan, S. E. (1984). *Language Assessment Scales*. Monterey, CA: CTB/McGraw-Hill.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52(3), 219-232.
- Deno, S. L., & Fuchs, L. S. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children*, 19(8), 1-16.
- Duncan, S. C., & Duncan, T. E. (1994). Modeling incomplete longitudinal substance use data using latent variable growth curve methodology. *Multivariate Behavioral Research*, 29, 313-338. doi:10.1207/s15327906mbr2904_1
- Elliott, J., Lee, S. W., & Tollefson, N. (2001). A reliability and validity study of the Dynamic Indicators of Basic Early Literacy Skills-Modified. *School Psychology Review*, 30, 33-49.
- Francis, D. J., Shaywitz, S. E., Stuebing, K. K., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology*, 88, 1-37. doi:10.1037/0022-0663.88.1.3
- Fuchs, L. S., & Deno, S. L. (1994). Must instructionally useful performance assessment be based in the curriculum?. *Exceptional Children*, 61, 15-24.
- Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children*, 57(6), 488-499.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect?. *School Psychology Review*, 22, 27-48.
- Good, R. H., & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), *Advanced application of Curriculum-Based Measurement* (pp. 61-88). New York: Guilford Press.
- Greenberg, E., Macias, R. E., Rhodes, D., & Chan, T. (2001). English literacy and language minorities in the United States. *Education Statistics Quarterly*, 3(4), 73-75.
- Habendank-Stewart, L., & Kaminski, R. (2002). Best practices in curriculum-based evaluation. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV*. Bethesda, MD: The National Association of School Psychologists.
- Harcourt Brace & Company (1997a). *Stanford Achievement Test Series - Ninth Edition: Spring norms book*. San Antonio, TX: Harcourt Brace & Company.
- Harcourt Brace & Company (1997b). *Stanford Achievement Test Series - Ninth Edition: Technical data report*. San Antonio, TX: Harcourt Brace & Company.
- Hasbrouck, J. E., & Tindal, G. (1992). Curriculum-based oral reading fluency norms for students in grades 2 through 5. *Teaching Exceptional Children*, 24(3), 41-44.
- Hintze, J. M., Callahan, J. E., Matthews, W. J., Williams, S. A., & Tobin, K. G. (2002). Oral reading fluency and prediction of reading comprehension in African American and Caucasian elementary school children. *School Psychology Review*, 31(4), 540-553.
- Howell, K. W., Kurns, S., & Antil, L. (2002). Best practices in curriculum-based evaluation. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp.753-770). Bethesda, MD: The National Association of School Psychologists.
- Hu, L.-Z., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new

- alternatives. *Structural Equation Modeling*, 6, 1-55. doi: 10.1080/10705519909540118
- Jimerson, S. R. (1997). *ORAL-J: The administration and technical manual*. Available from S. R. Jimerson, University of California, Santa Barbara, California, 93106-9490.
- Jimerson, S. R. (2000). *ORAL-J: The administration and technical manual*. Available from S. R. Jimerson, University of California, Santa Barbara, California, 93106-9490.
- Klein, J., & Jimerson, S. R. (in press). Examining ethnic, gender, and language bias of oral reading fluency scores among Caucasian and Hispanic students. *School Psychology Review*.
- Kranzler, J. H., Miller, M. D., & Jordan, L. (1999). An examination of racial/ethnic and gender bias on Curriculum-Based Measurement of reading. *School Psychology Quarterly*, 14, 327-342. doi:10.1037/h0089012
- Marston, D. B. (1987). The effectiveness of special education: A time series analysis of reading performance in regular and special education settings. *Journal of Special Education*, 21, 13-26.
- Marston, D., & Magnusson, D. (1985). Implementing curriculum-based measurement in special and regular education settings. *Exceptional Children*, 52, 176-266. doi:10.1177/002246698802100405
- McArdle, J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, 58, 110-133. doi:10.2307/1130295
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107-122. doi:10.1007/BF02294746
- Powell-Smith, K. A., & Bradley-Klug, K. L. (2001). Another look at the "c" in cbm: Does it really matter if curriculum-based measurement reading probes are curriculum-based?. *Psychology in the Schools*, 38(4), 299-312. doi:10.1002/pits.1020
- Schafer, J., & Olsen, M. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, 33, 545-571. doi:10.1207/s15327906mbr3304_5
- Shinn, M. R. (Ed.) (1998). *Advanced Applications of Curriculum-Based Measurement*. New York: The Guilford Press.
- Shinn, M. R. (Ed.) (1989). *Curriculum-Based Measurement: Assessing Special Children*. New York: The Guilford Press.
- Sibley, D., Biwer, D., & Hesch, A. (April, 2001). *Establishing curriculum-based measurement oral reading performance standards to predict success on local and state tests of reading achievement*. Paper presented at the Annual Meeting of the National Association of School Psychologists, Washington, DC.
- Stage, S. A. (2001). Program evaluation using hierarchical linear modeling with curriculum-based measurement reading probes. *School Psychology Quarterly*, 16, 91-112.
- Stage, S. A., & Jacobsen, D. M. (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review*, 30, 407-419. doi:10.1521/scpq.16.1.91.19159
- Stage, S. A., Sheppard, J., Davidson, M., & Browning, M. (2001). Prediction of first grader's growth in oral reading using their kindergarten letter-naming and letter-sound fluency. *Journal of School Psychology*, 39, 225-237. doi:10.1016/S0022-4405(01)00065-6
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360-407. doi:10.1598/RRQ.21.4.1
- US. Department of Education, National Center for Education Statistics (2003). *Status and trends in the education of Hispanics* (NCES 2003-08). Washington, DC: Author.
- Tindal, G. (1993). A review of curriculum-based procedures on nine assessment components. In J. J. Kramer (Ed.), *Curriculum-based measurement* (pp.25-64). Lincoln, NE: Buros Institute of Mental Measurements.
- Whitehurst, G. J., & Fischel, J. E. (2000). Reading and language impairments in conditions of poverty. In D. V. M. Bishop & L. B. Leonard (Eds.), *Speech and language impairments in children: Causes, characteristics, intervention and outcome* (pp. 53-71). Hove, England: Psychology Press.