CRESST REPORT 779

Jinok Kim
Joan L. Herman

WHEN TO EXIT ELL STUDENTS: MONITORING SUCCESS AND FAILURE IN MAINSTREAM CLASSROOMS AFTER ELLS' RECLASSIFICATION

DECEMBER, 2010

The National Center for Research on Evaluation, Standards, and Student Testing

Graduate School of Education & Information Sciences
UCLA | University of California, Los Angeles

**When to Exit ELL Students:
Monitoring Subsequent Success and Failure in
Mainstream Classrooms after ELLs' Reclassification**

CRESST Report 779

Jinok Kim & Joan L. Herman
CRESST/ University of California, Los Angeles

December, 2010

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Bldg., Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported under the National Research and Development Centers, PR/Award Number R305A090581, as administered by the U.S. Department of Education, Institute of Education Sciences.

To cite from this report, please use the following as your APA reference:
Kim, J., & Herman, J. L. (2010). *When to exit ELL students: monitoring subsequent success and failure in mainstream classrooms after ELLs' reclassification.* (CRESST Report 779). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

TABLE OF CONTENTS

WHEN TO EXIT ELL STUDENTS:

MONITORING SUBSEQUENT SUCCESS AND FAILURE IN MAINSTREAM CLASSROOMS AFTER ELLS' RECLASSIFICATION

ANNUAL REPORT

Jinok Kim & Joan L. Herman
CRESST/ University of California, Los Angeles

Abstract

In English Language Learners' (ELLs) reclassification, the tension between assuring sufficient English language proficiency (ELP) in mainstream classrooms and avoiding potential negative consequences of protracted ELL status creates an essential dilemma. This study assesses the validity of existing systems in terms of gross consequences of reclassification. We examine the subsequent academic success of reclassified ELLs in mainstream classrooms, using statewide individual-level data merged from Grades 3 to 8 in a local control state. Drawing on some recent advances in growth modeling techniques, we control for students' performance levels prior to reclassification in examining post-reclassification growth rates. The study found that ELL students tend to make a smooth transition upon their reclassification and keep pace in mainstream classrooms. This indicates that existing reclassification decisions are, in general, supportive of ELL students' subsequent learning, with a caution that such finding should be tempered by a great extent of heterogeneity in subsequent learning. Results on one main component of reclassification criteria, ELP levels upon reclassification, suggest that protracted ELL status due to too stringent ELP criteria may not be useful or even be detrimental to ELLs' learning in mainstream classrooms.

Introduction

Reclassification is a key milestone for English Language Learners (ELLs) in what they experience in schools. Reclassification is the point when students are expected to fully function in mainstream classrooms, without any further special English language development (ELD), instructional services, or assessment accommodations. Consequently, faulty decisions about their readiness may seriously hamper future learning. Strong claims have been made, for example, that prematurely exiting ELLs out of ELD programs can have detrimental effects (Cummins, 1980; 1981). At the same time, other researchers have raised concerns about the potential adverse consequences to ELL students who remain in that status for extended periods of time. ELL status in secondary schools may functionally mean less access to the math and science classes that are required for high school graduation and admission to post-secondary education (seeParrish, Perez, Merickel, & Linquanti, 2006). The cumulative effects of diminished access to academic coursework over time can be significant, potentially preventing ELLs from entering post-secondary education (Callahan, 2005; Harklau, 2002). Moreover, negative affective consequences of ELL status during adolescence have been noted (Gándara, Gutierrez, & O'Hara, 2001; Maxwell-Jolly, Gándara, & Méndez Benavídez, 2007), and research shows a strong relationship between ELL status and school dropouts (Silver, Saunders, & Zarate, 2008; Watt & Roessingh, 1994).

The tension between assuring that students have sufficient ELP to be successful in mainstream classrooms and avoiding the potential negative consequences of protracted ELL status creates an essential dilemma in determining the optimal time for ELL reclassification. Lacking firm evidence to resolve the dilemma, states and local schools show variation in their reclassification criteria and procedures; furthermore, inconsistencies, inadequacies, and ambiguities have repeatedly been reported both within states (see Linquanti, 2001; Parrish et al., 2006) and between states (see statewide practice review by Wolf et al., 2008). The validity of existing criteria and procedures lack an empirical base; reclassification policies and practices are formulated and implemented with little knowledge of the factors that may influence their success.

The current research project plans to provide much needed evidence to begin to resolve these current dilemmas. Our study takes a particular view on the meaning of effective reclassification: it can and should be judged by its consequences. Therefore, optimal reclassification systems, in our opinion, are those that maximize subsequent success while minimize failures. Reclassified students should be ready to benefit from mainstream classrooms without the provision of special services. Students' ability to benefit should be evidenced in subsequent outcomes such as academic performance on state tests of reading and mathematics, passing high school exit exams, and persisting in school.

Thus, reclassification systems can be considered valid or even optimal, under certain circumstances, if they are associated with increases in positive consequences and reductions in negative consequences for student educational outcomes. For example, there is evidence of validity to the given reclassification systems and/or criteria if reclassified students learn more rapidly after reclassification than before; if they catch up to non-ELL students in mainstream classrooms and close gaps in the following years; or if they learn as well as non-ELL students under similar conditions. If those particular reclassification criteria are related to a reduction in subsequent dropout rates, it could also provide support for those criteria. Based on such a perspective, we provide a framework in which (a) the validity of existing systems can be examined and assessed in terms of gross consequences of reclassification—or the subsequent academic success or failure of reclassified ELLs in mainstream classrooms; and (b) differences in reclassification criteria as well as in various student, school, and district factors related to differences in relative success in promoting subsequent student achievement.

The larger study— of which the current work is a part—purposely includes two states in order to reflect the range of reclassification policies and practices that are currently implemented. Using statewide data from multiple states, we identify student groups by ELL status over multiple grades, especially ELL students who are reclassified at Grades 4, 5, or 6. We apply growth modeling techniques that are suitable when studying data that have time series (see Diggle, Liang, & Zeger, 1994; Raudenbush & Bryk, 2002, Chapter 6; Singer & Willet, 2003, Chapter 3). By longitudinally monitoring ELLs' academic achievement over the years before and after reclassification, this study will highlight the academic growth patterns over time of students reclassified in various grades, which are subgroups that have been understudied. In comparing intact groups (e.g., comparing reclassified ELL students or other ELL students), this study draws on the strength of some of the recent advances in growth modeling techniques that allows for regressions among latent variables or growth parameters (Choi & Seltzer, 2010; Muthen & Curran, 1997; Seltzer, Choi, & Thum, 2003). By holding constant prior status in examining subsequent growth rate, the method increases the comparability of intact groups in their growth patterns.

The ways and the degree to which reclassified ELL students benefit from mainstream classrooms can depend on various factors. These factors may include reclassification criteria used, student characteristics, and practices around reclassification— to name but a few. This research project goes beyond the average differences but examines for whom, under which criteria, and under which settings, reclassified ELL students receive greater benefits and experience more success. . In addition to examining student and district factors as well as individual-level reclassification criteria used, which are available from extant state data in the present study (see the next section for an overview), the larger study of which the current study is a part also includes a qualitative component. In addition to the extant state data, we will collect detailed information about existing reclassification criteria, and ELL programs and services from states, local districts, and schools. Such qualitative data will be incorporated into our quantitative analysis. This will enable a more in-depth study of features of reclassification criteria and/or the ELL programs and services that are associated with districts or schools that foster higher versus lower growth.

The Present Study

In collaboration with one state's (State A, henceforth) Department of Education, the study reported here used statewide individual-level data merged across multiple academic years to track two cohorts of students over a six-year period. The data include variables such as ELD performance (for years prior to reclassification); academic achievement based on state content assessment, dropouts; and other demographics (e.g., eligibility for free or reduced lunch, ethnicity, homeless status) for six years for both ELL and non-ELL students. For ELL students, variables such as ELP level and years of ELL status were obtained for the same time period. The state annual assessments are vertically equated and thus the assessment scales are comparable across grades.

These longitudinal data include one cohort starting at each of the elementary and secondary school levels. For the first, younger cohort, we obtained six years of longitudinal data on the statewide cohort that started in Grade 3 in 2003–2004, which enabled us to track them through Grade 8 (in the year 2008–2009). As for the second, older cohort, we obtained six years of longitudinal data for students who were in Grade 6 in the same year (i.e., 2003–2004) which enabled us to track them to Grade 11 (in the year 2008–2009). The present study examines academic outcomes in reading and math for the younger cohort only, from one state (State A). Ensuing reports will include studies of the older cohort from State A and of the other state, as well as studies of other outcomes such as dropouts for both states.

The present study focuses on ELL students who are reclassified at Grades 4, 5, and 6, because (a) students reclassified in those grades can be identified with more certainty from State A's data which began tracking students from Grade 3; and (b) those are the ELL students who get reclassified before they finish elementary school or right when they finish the first year of middle school. Since these are the ELL students who are not initially fluent in English but are reclassified before they become long-term ELLs, they form one of the critical sub-populations for the study of the reclassification of ELL students.

Information about ELLs who were reclassified at Grades 4, 5, and 6 was not immediately available from the state data. The state assessment data do have information about students' ELL status for every academic year, but merging ELL status across the six years yielded too many patterns to be easily understood. In this paper, we refer to students' ELL status for the six years as ELL status profiles. We describe ELL status profiles of one entire cohort of ELLs in State A as well as the procedures we used to identify ELLs reclassified at Grades 4, 5, and 6.

Not only was the identification of reclassified ELLs from the data complex but also practices around reclassification may have varied within the state. According to a federal guideline about the reclassification of ELLs, reclassified ELL students should be monitored in their ELP levels for two subsequent years. In State A, students are monitored for two years when students' ELP labels change from Limited English Proficient (LEP) to Fully English Proficient (FEP; State A's Department of Education, 2007). However, a change from LEP to FEP does not necessarily mean "reclassification" because some schools or districts may continue a similar level of assistance in terms of ELD. Since State A is a state with local controls, practices regarding ELL students including reclassification are decided at the district or school levels. Whether students are mostly mainstreamed or received as much ELD assistance as before is determined at a local level (personal communication with state personnel, 2009). Such varying practices may make it difficult to know the exact timing of reclassification from any data system.

In addition, the reclassification criteria are also at the discretion of local school districts. The main guidelines from the state regarding the criteria for ELL reclassification are: (a) reaching the overall level of 5 (highest level) in the state ELP assessment and (b) reaching the level of *Partially Proficient* in the English version of the state assessment in reading and writing (State A's Department of Education, 2007). Districts are also advised to use multiple informational sources in their decision-making process, including the results of State A's statewide ELD and content assessments (State A's Department of Education, 2007; see also Escamilla, Mahon, Riley-Bernal, & Rutledge, 2001). Based on personal communication with state personnel, districts usually follow the guidelines and suggestions made by the state. However, in the end, the specific criteria used for reclassifying ELL students are at the discretion of local districts or schools. We utilize state longitudinal data and trace back which criteria may have been used for individual students. As it will be seen, the results show that districts and schools made more exceptions than precisely following the state guidelines.

The present study uses growth modeling techniques (Diggle, Liang, & Zeger, 1994; Raudenbush & Bryk, 2002; Singer & Willet, 2003) and tracks the academic progression of students in reading (Grade 3 through Grade 8) and over four grades in math (Grades 5 through 8). As we focus on the consequences of the reclassification system in order to provide its validity evidence, the primary outcome of interest is the academic growth of ELL students after reclassification. As noted, this present study focuses on students who are reclassified at Grades 4, 5, and 6. The primary outcome of the study, the academic growth after reclassification, corresponds to students' growth in Grades 4 through 8 for those reclassified at Grade 4; Grades 5 through 8 for those reclassified at Grade 5; and Grades 6 to 8 for those reclassified at Grade 6. For each time period, a student's performance level at the first year is the performance level right before the year he or she exits or the year upon exiting. We approximate growth in these three time periods by examining a student's growth from Grades 5 to 8. Therefore, our primary outcome is academic growth in Grades 5-8, which we refer to as *middle-school growth* or as *post-reclassification growth* in this study.

Assessing post-reclassification growth can be thought of in two different ways. Post-reclassification growth can be compared to pre-reclassification growth of the same individuals, or, alternatively, post-reclassification growth can be compared to post-reclassification growth of other individuals. The first way is related to the idea of interrupted time series (ITS) design (Shadish, Cook, & Campbell, 2002, Chapter 6; Campbell & Stanley, 1966), in which repeated observations of the same group of individuals are compared before and after the introduction of treatment. Abrupt changes in time trend upon the introduction of a treatment can be considered as the effect of treatments. The ITS design was not employed in this study for various reasons. For the most part, although we implicitly view existing reclassification and the criteria underlying it as an intervention, it is not an intervention that is meant to make abrupt changes, unlike other more typical interventions. For example, if reclassified students have caught up to their non-ELL peers prior to reclassification, and after reclassification they continue such trend over time, that would imply a smooth transition to mainstream classrooms and therefore support validity of the reclassification decision that was made. There were also technical and other issues for why we did not employ ITS design, which we revisit in later sections (see methods and discussion sections).

The second way, which we employed in this study, is to compare post-reclassification growth of reclassified ELLs with growth of other students in the same period of time. Specifically, we compare our target groups (i.e., ELL students reclassified at Grades 4, 5, and 6) with non-ELL students and other ELLs who were not reclassified during the above three grades. Since we deal with groups with different characteristics and different performance levels before the target period, it may not be meaningful to compare post-reclassification growth to see if reclassified ELLs grow more or less rapidly than they would have grown otherwise. To alleviate such difficulty arising from comparing groups with different characteristics, we control for – or hold constant – students' performance status before reclassification (i.e., performance at Grade 5). We apply recent advances in growth modeling techniques to the data, in which growth parameters can be regressed on the other growth parameters in the model, as well as other covariates or predictors of interest (Choi & Seltzer, 2010; Muthen & Curran, 1997; Seltzer, Choi, & Thum, 2003).

If there is appreciable variability in how students grow in academics subsequent to reclassification, it is important to investigate for whom, under which criteria, and under which settings, reclassified ELL students benefit more and their success is more enhanced. In addition to examining the validity of existing ELL reclassification systems by assessing the expected post-reclassification growth, the present study goes beyond average growth and explores differences in post-reclassification growth across individuals and districts. We include various information on students and districts in growth models to see how differences in post-reclassification growth relate to differences in student characteristics, reclassification criteria, or district membership.

Specifically, for the entire cohort of students, including reclassified ELLs, non-ELLs, and other ELLs, we examine the following two questions:

1. How does the estimated average middle-school academic growth of reclassified ELL students (i.e., ELL students reclassified at Grades 4, 5, and 6) compare to the average middle-school academic growth of non-ELL students or other ELL students?
2. To what extent do the estimated students' growth trajectories vary across individual students?

Then, we zero in on reclassified ELLs and examine the following to see for whom and under which settings reclassified ELLs tend to show more enhanced academic growth subsequent to reclassification:

3. What are the characteristics of reclassified ELLs that are associated with greater subsequent academic success?
4. To what extent do the estimated growth trajectories of reclassified ELLs vary across 38 districts?
5. How do differences in ELP levels relate to differences in subsequent academic success?

Data and Study Sample

State A's Longitudinal Data and Descriptive Statistics

Annual state assessments for all students in the state, annual ELP assessments for all current and monitored ELL students, and student demographics comprise the main part of the state data. State A's annual, standards-based assessment has been in place since 1997 and is administered to 98% of all students in Grades 3–10 in reading, mathematics, and writing. The annual assessment is a vertically equated test system, which allows for meaningful comparisons across grade levels (State A's Department of Education, 2007, Part 6), and its validity and reliability has been documented (State A's Department of Education, 2007, Part 8).

State A also has an ELP test that is administered to ELLs, kindergarten through Grade 12, and the grade spans for each test form are as follows: K, 1, 2, 3–5, 6–8, and 9–12. This ELP test is designed to assess ELP levels in four modalities: speaking, listening, reading, and writing. It produces scale scores and proficiency levels from 1 (*beginning*) to 5 (*advanced*) based on each of these modalities. Additionally, the ELP test produces an overall measure of comprehension (derived from the listening and reading components) and an oral language composite (derived from the listening and speaking components).

Table 1 presents percentages for student demographics available from the state data for the entire population in the cohort as well as percentages by ELL status. Students are considered ELL students if they were placed in an ELL program at any time during the six-year data. That is, for the younger cohort, if students were placed in ELL programs in any grade during Grades 3–8, they were considered ELL students. More details about classifications of ELL students based on the state longitudinal data set are discussed in the next section.

Table 1

Demographic Information for All Students (the Cohort of Third Graders in the Year 2003–2004), and by ELL status, from State A's Data System

Demographics	All students (*n*=55,033)	ELL (*n*=7,667)	Non-ELL (*n*=47,366)
Native American	1.2%	1.0%	1.0%
Asian/Pacific Islander	3.4%	7.0%	3.0%
Black	6.1%	1.0%	7.0%
Hispanic	27.4%	87.0%	18.0%
White	61.9 %	5.0%	71.0%
Disability status	10.2%	11.0%	10.0%
Migrant status	0.6%	3.0%	0.0%
Immigrant status	0.5%	3.0%	0.0%
Economically disadvantaged	36.5%	81.0%	29.0%
Homeless status	1.0%	2.0%	1.0%

Table 1 shows that, based on the student population members who were third graders during the 2003–2004 year, State A's student population in public schools consists of 62% White; 27% Hispanic; 6% African American; 3% Asian and Pacific Islanders; and 1% Native Americans. Ten percent of the student population is identified as students with disabilities; while 37% of the students are economically disadvantaged and about 1% of the students are reported to be homeless.

Table 1 also shows that the majority of State A's ELL population is Hispanic (87%). Other categories include Asian and Pacific Islanders (7%) and White (5%). The ELL student population is also disproportionately more economically disadvantaged (81% versus 29% in the non-ELL population) and has more homeless students (2% versus 1% in the non-ELL population). The findings in demographics are consistent with general trends with other cohorts in the state and with other states (Kim & Herman, 2008; Parrish et al., 2006).

Table 2 presents descriptive statistics of reading achievement scores in the annual state assessment for all students in the cohort, and Table 3 presents the reading scores by ELL status. The average score of ELL students at Grade 6 is similar to the average score of non-ELL students in Grade 3, which represents a significant achievement gap between ELL and non-ELL students. The standard deviations (*SD*s) of scores tend to decrease as the grade levels increase both for ELL and non-ELL students. For the entire cohort of students, the *SD* is greatest at Grade 3 (meaning that student scores are most dissimilar in Grade 3); the *SD*s decrease about 10 points for the next three years, and decrease again respectively at Grades 8 and 9. This may imply different scaling compression in different grades. In addition, it is notable that the average scores increase rather modestly over grades, with the increment smaller in later grades, while the minimum scores increase by a larger amount more similarly in all grades. This may further suggest that the scaling compression may not only be different in different grades but also in different ranges of scores. Such scaling issues may complicate interpretation of findings when we track students over time, as will be seen.

Table 2

Descriptive Statistics of Reading Scores in the State Annual Assessment

Reading Score	*N*	*M*	*SD*	Min	Max
Grade 3	54236	563.59	75.10	150	795
Grade 4	50896	586.95	65.76	180	940
Grade 5	49223	614.86	67.00	220	906
Grade 6	47300	626.21	67.57	260	839
Grade 7	45980	640.66	62.00	300	891
Grade 8	45080	649.91	57.14	330	927

Table 3

Descriptive Statistics of Reading Scores in the State Annual Assessment, by ELL Status

Reading score	*N*	*M*	*SD*	Min	Max
ELLs					
Grade 3	7106	504.98	71.30	150	795
Grade 4	6920	529.51	66.53	180	733
Grade 5	6762	560.27	69.14	220	772
Grade 6	6456	571.17	70.03	260	788
Grade 7	6157	594.81	58.94	300	821
Grade 8	5984	608.39	54.68	330	784
Non-ELLs					
Grade 3	47130	572.43	71.61	150	795
Grade 4	43976	595.99	60.88	180	940
Grade 5	42461	623.55	62.39	220	906
Grade 6	40844	634.90	62.91	260	839
Grade 7	39823	647.75	59.38	300	891
Grade 8	39096	656.26	54.79	330	927

Table 4 presents descriptive statistics of math achievement scores in the annual state assessment for all students in the cohort, and Table 5 presents the math scores by ELL status. The average score of ELL students at Grade 6 is lower than the average score of non-ELL students at Grade 4, which represents a significant achievement gap between ELL and non-ELL students. The *SD*s of scores are similar from Grades 4–7, but decreases by about 10 points in Grade 8 for both ELL and non-ELL students. In addition, it is notable that the average scores increase rather modestly over grades, except between Grades 4 and 5, whereas the minimum scores increase by a larger amount. As in reading, this may suggest that the scaling compression may be different in different grades as well as in different range of scores.

Table 4

Descriptive Statistics of Math Scores in the State Annual Assessment

Math score	*N*	*M*	*SD*	Min	Max
Grade 4	50922	482.77	74.17	180	780
Grade 5	49224	521.53	73.67	220	800
Grade 6	47373	539.39	75.99	240	830
Grade 7	46040	550.36	72.29	280	860
Grade 8	45107	575.20	61.45	310	890

Table 5

Descriptive Statistics of Reading Scores in the State Annual Assessment, by ELL Status

Math score	*N*	*M*	*SD*	Min	Max
ELLs					
Grade 4	6914	428.09	68.45	180	720
Grade 5	6768	472.81	65.52	220	732
Grade 6	6472	488.84	70.70	240	752
Grade 7	6158	503.46	67.54	280	761
Grade 8	5989	537.74	57.65	310	723
Non-ELLs					
Grade 4	44008	491.36	71.31	180	780
Grade 5	42456	529.30	71.89	220	800
Grade 6	40901	547.39	73.68	240	830
Grade 7	39882	557.60	70.26	280	860
Grade 8	39118	580.93	59.98	310	890

Profiles of ELL Status Based on State A's Longitudinal Data

When the state annual data system is merged longitudinally, it contains rich information and may reveal complexities in practices. This is the case when we identify ELL students and reclassified ELL students during the six-year span. We relied on two variables available from state data sets that point to ELL status most directly: placement in bilingual and English as a Second Language (ESL) programs. We need further information regarding how bilingual and ESL programs were defined, and whether a certain selection method by schools or districts was used to decide between the two programs. For the current purpose of identifying ELL profiles, we treated both variables equivalently. Descriptive statistics for the two variables show that a majority of ELL students were placed in ESL programs compared to bilingual programs. Table 6 also displays the codes that were used for both variables (i.e., placement in either the bilingual or ESL program).

Table 6

State Data Coding Scheme for Two ELL Status Variables: Bilingual and ESL

Value	Variable
0	No (not enrolled in any program)
1	Yes (enrolled in a program)
2	Redesignated program (monitored Year 1)
3	Redesignated program (monitored Year 2)
4	Exited program (Year 3+)
5	Not in the program from parent's choice

In order to capture ELL status profiles during the six year span, we created one variable using six ESL status variables for each year. We basically connected the values of the six ESL status variables in the chronological order, with the values of each ESL status variable listed in Table 6. For example, the ELL status profile with a value of "000000" would indicate Non-ELL students, the profile value "111111" would represent long-term ELLs because they have been enrolled in a program for all six years until Grade 8, and "111234" represents ELL students who have exited a program when they advance to Grade 6.

We examined this new variable that shows ESL status profiles for the entire cohort of ELL students. The results show that so many profiles (1,753 profiles) emerged that it would be difficult to identify different groups such as reclassified ELLs and other ELLs just based on the profiles. For example, among the 1,753 profiles, 1,042 profiles have a frequency of 1, which means that each of those ESL profiles is unique to one student.

Table 7 shows excerpts from the frequency table of the variable of ESL status profile – using just the ESL (not the Bilingual) variable – selecting only 30 of the profiles that have the most frequencies from the entire list of 1,753 profiles. From Table 7, one can see that the "0" (which indicates Non-ELL, see Table 6) and missing values (indicated by the asterisk [*] symbol) may further complicate the profiles in some cases. Missing values or 0 values may be recorded for some ELL students who may not have been assigned to an ESL program in that year for various issues, or for some reclassified ELL students. For example, for some reclassified ELL students who appear to speak English fluently, a person who is marking the ELL status of students for the data system (e.g., a school ESL teacher) may not have known the history that these students used to be ELLs. In such settings, the teacher may have just coded them as Non-ELLs (value of 0) or even may not record the value.

Table 7

Excerpts from the Frequency Table of the ESL Profile Variable

Serial number	ESL profile	N	Percent (%)
1	000000[a]	35771	65.00
2	0******[b]	2471	4.49
3	00****	1477	2.68
4	000***	1447	2.63
5	00000*	1259	2.29
6	0000**	1134	2.06
7	111111	784	1.42
8	0000*0	374	0.68
9	1*****	292	0.53
10	000*00	280	0.51
11	0*0000	269	0.49
12	123444	231	0.42
13	00*000	230	0.42
14	040000	206	0.37
15	111112	197	0.36
16	011111	188	0.34
17	000111	147	0.27
18	234444	145	0.26
19	000**0	143	0.26
20	001111	140	0.25
21	11****	140	0.25
22	111234	137	0.25
23	0**000	128	0.23
24	000440	128	0.23
25	00**00	121	0.22
26	000554	120	0.22
27	111123	109	0.20
28	120444	105	0.20
29	0***00	100	0.18
30	010344	99	0.18

[a]See Table 6 for explanations for the coded values. [b]An asterisk (*) indicates missing data for that year.

Many profiles from the 30 most frequent profiles seem to be Non-ELLs, as can be expected. Not surprisingly, the first profile that comprises 65% of the entire student cohort is Non-ELL. Also, many other profiles, such as serial numbers 2, 3, 4, 5, 6, 8, 10, 11, 13, may also be Non-ELLs since missing values may have arisen in some of the years due to the reporting system or other issues, but in all the other years they are recorded to be non-ELLs. We consider these profiles as non-ELL profiles in this study, unless those students are indicated as ELLs by the other ELL status variable (based on a bilingual program). More details follow in the next section.

Among ELL profiles, the most frequent one is “111111” with the frequency of 784 (serial number 7), which would represent long-term ELLs. The second most frequent ELL profile is “1…” (Serial Number 9), which means that in Grade 3 these students attended the ESL program, but there are missing ESL values from Grades 4 through 8. Some of these students may have exited from the program, and others may have been unavailable to follow up (e.g., moved out of state). The third most frequent profile is “123444” (serial number 12) with the frequency of 231, which means that these ELL students attended ESL programs in Grade 3, were monitored in Grades 4 and 5, and exited during Grade 6. The fourth most frequent profile is “040000” (serial number 14) with the frequency of 206. Some of these students may be initially fluent in English; in other words, although they speak another language at home, they have been fully fluent in English in school. One other profile that may include initially English fluent students in Table 7 is “000440” (serial number 24, frequency of 128). The fifth most frequent profile is “111112” (serial number 27), comprised of students who were in the ESL program from Grade 3 all the way to Grade 7, and started to be monitored in Grade 8.

The next three most frequent profiles include students who were not initially a part of any ESL program, but then in later grades they attended ESL programs: “011111” (serial number 16, frequency = 188); “000111” (serial number 17, frequency = 147); and “001111” (serial number 20, frequency = 140). Note that students with these profiles include students who started with bilingual programs and later attended ESL programs. In such cases, the student have a value of Bilingual variable of “1”, which means attending a bilingual program for the earlier grade(s), whereas in the same year(s) he or she shows the ESL variable value of “0,” which indicates not attending an ESL program. Although such information is not shown here, we also utilize profiles from the Bilingual program variable, and take such cases into account in identifying different groups based on ELL status profiles (see the next section).

Students with “000554” (serial number 26, frequency = 120) include students who did not attend any ESL programs during elementary school, but exited the program during Grades 6 and 7 due to parent(s)’ decision to not continue their child in the program. Students with “111123” (serial number 27) profile include students who were placed in the ESL program until Grade 6 and started to be monitored at Grade 7. The profiles of "120444" or "010344" (serial numbers 28 and 30, respectively) may not make sense if we take the value "0" literally as non-ELL students. However, as noted earlier, it is probably the case that in the grades where students did not attend an ESL program they may have been marked as non-ELLs (i.e., the value of "0") instead of monitor or exit.

Identifying Reclassified ELLs Based on State A's Longitudinal Data

The profiles of ELL status during the six-year span described in the previous section illustrate only part of the complexity that researchers may encounter in understanding ELL status over time. Note that we covered only the 30 most frequent ESL profiles among more than 1,700 profiles, and that the previously described profiles are based on only one of the two ELL status variables (i.e., the ESL program variable, but not the Bilingual program variable).

Too many profiles from the state longitudinal data exist, which may be in part due to insufficient consistency with the coding by different people in different settings, or from other logistical reasons. Such complexity may also come from the ELL students. There may be students who move often and change schools, which may result in difficulty in following up with precise information in the immediate year. Availability of ESL programs or different types of programs may change for various reasons, such as students advancing to different school levels, and changes in programs and policies at local districts or schools.

In identifying reclassified ELLs from state longitudinal data, the extent of the complexity increases when we consider the variability of practices regarding ELL reclassification across local districts and schools. As reviewed in a previous section, State A monitors students for two years after ELL students reach FEP status, which may or may not be full redesignation. In State A, when an ELL student is in a monitored status (the ELL status is coded as "2" for the first year monitoring and as "3" for the second year monitoring; see Table 6), the student may still receive ELL assistance, or may have been fully mainstreamed, depending on the local policy. Therefore, this is slightly different from federal guidelines that suggest two years of monitoring after ELLs' exit from the ELL program. Furthermore, as will be seen, investigating the longitudinal data set revealed even more flexibility than the situations discussed above. For example, many ELL students can be placed in monitored status before they reach FEP.

All evidence indicates complexity in practices around reclassifying or exiting ELL students, and one might say that the definition of reclassification may be fairly ambiguous in such settings. This means that, depending on how we operationalize the definition of reclassification, we might end up with different groups of ELLs. Table 8 presents a scheme we developed and used in this report to identify current ELLs and ELLs who were reclassified during Grades 4, 5, and 6.

Table 8

Definitions of ELLs and Reclassified ELLs Based on State Longitudinal Data

Category	Definition
ELL	Students who were placed in any ELL program in any grade from Grade 3 to Grade 8
ELL broad	Students who were placed in any ELL program in any grade from Grade 3 to Grade 8 (i.e., the above "ELL" category); and students who were marked as monitor or exit status in any grade from Grade 3 to Grade 8
Reclassified during Grade 4	Among students who were placed in any ELL program in Grade 3, students whose ELP level was FEP in the previous year (Grade 3); and/or students who showed monitor or exit status in Grade 4 (coding values of 2, 3, 4 in Bilingual and ESL variables)
Reclassified during Grade 5	Among students who were placed in any ELL program in any year from Grade 3 to Grade 4, students whose ELP level was FEP in the previous year (Grade 4); and/or students who showed monitor or exit status in Grade 5 (coding values of 2, 3, 4 in Bilingual and ESL variables)
Reclassified during Grade 6	Among students who were placed in any ELL program in any grade from Grade 3 to Grade 5, students whose ELP level was FEP in the previous year (Grade 5); and/or students who showed monitor or exit status in Grade 6 (coding values of 2, 3, 4 in Bilingual and ESL variables)

Using SAS software (SAS Institute Inc., 2002–2005), we implemented programs to select students in each category as defined in Table 8. The "ELL" category groups ELL students in a narrow sense— by including only the students who were ELLs in any grade during the period that we examined (six years from Grade 3 to Grade 8). On the other hand, the "ELL broad" category groups ELL students in a broad sense by including students in the "ELL" category as a subset in addition to including students who have been in monitor status during the six-year period. The additional set of students includes students who were redesignated during Grade 2 and students who spoke languages other than English at home, but initially spoke fluent English upon entering school.

The "Reclassified during Grade 4," "Reclassified during Grade 5," and "Reclassified during Grade 6" categories are for reclassified students. The three reclassified categories are mutually exclusive, and all three categories are a subset of the "ELL" category. Students are defined to be reclassified in the respective grades if (a) their language classification in the previous year was FEP (this is suggested from the state's policy); and (b) their ELL status is monitor or exit in the year of redesignation. More students were reclassified due to criteria (b) than (a), which means that a large proportion of students are redesignated before students reach the FEP level in their English language classification.

The five aforementioned categories were developed through an iterative process. The iterations were necessary mainly because of two ELL status variables and profiles based on each. Also, given the large number of ELL status profiles (more than 1,700 based on one of the two status variables only), some were most likely to fall into one of the reclassified categories when checked individually but cannot be captured to be in the categories in initial iterations of programs. We applied one set of programs to the data, checked ELL profiles from the resulting categories, and refined programs to capture profiles that were left out. We continued this process for each category in multiple iterations until all profiles were accounted for.

Study Sample

State A is a local control state that consists of more than 180 districts. The districts have a considerable range in various demographic characteristics. In terms of enrollment size, some of the districts are large, with thousands of students in a cohort, while most of the districts are very small, with less than a couple hundred students in a cohort. In addition, many districts do not have any ELL students enrolled in their districts, whereas in some districts, ELL students comprise more than 50% of the student population. A target population of this study is students who were reclassified during Grades 4, 5, and 6. Thus, for the purpose of this study, we included districts that have more than 20 ELL students enrolled, using our definition of ELLs (see Table 8). This resulted in only 38 districts in our study sample from the more than 180 districts in the state data set. Although this may seem to be a huge reduction at first glance, this does not mean much loss of information for the purpose of this study (e.g., studying ELL reclassification policies). A vast majority of the districts that were left out from our study sample had none to a minimal number of students in our target population (i.e., students who were reclassified during Grades 4, 5, and 6). Also, because these districts tend to be very small districts, the non-ELL population did not decrease much either. Overall, we retained 82% of the entire population and 94% of the ELL population. Thus, the reduction of sample helps our study focus on issues around ELL students more clearly with none or trivial cost in terms of generalization.

Table 9 displays demographics of the study sample for all students as well as ELL status. The demographic statistics of the study sample are almost identical to the statistics from the entire student sample, which confirms again that the study sample is representative of the states' entire population in the cohort.

Table 9

Descriptive Statistics of Demographic Information for the Study Sample

Demographics	All students (*n*=45,006)	ELL (*n*=7,198)	Non-ELL (*n*= 37,808)
Native American	1.1%	0.8%	1.2%
Asian/Pacific Islander	3.7%	6.6%	3.2%
Black	6.6%	1.2%	7.7%
Hispanic	29.5%	86.9%	18.6%
White	59.0%	4.5%	69.3%
Disability status	10.4%	11.3%	10.2%
Migrant status	0.6%	3.1%	0.1%
Immigrant status	0.6%	3.2%	0.1%
Economically disadvantaged	37.6%	81.0%	29.4%
Homeless status	1.2%	2.0%	1.0%

Table 10 shows the 38 districts (with pseudo-district IDs in the state data files that were available to the study) in our sample, as well as ELL and demographic characteristics of the districts. The districts are sorted by decreasing number of ELLs, which means that districts with more ELLs are shown in the upper rows. The 38 districts that were selected to be in the study sample tend to be districts with relatively larger enrollment sizes but the study sample still includes districts with enrollment sizes of 150 or less in the cohort. Several large districts have thousands of students. The percentages of ELLs in the districts vary widely from under 3% to 55%. Districts with a larger percentage of ELLs tend to be the districts with more proportion of Hispanic students (r = 0.65); ones with less proportion of White students (r = 0.66); ones with more migrant and immigrant students (r = 0.41 and 0.45, respectively); ones with more economically disadvantaged students (r = 0.27); and ones with more homeless students (r = 0.26).

Table 10

The 38 Districts in the Study Sample and ELL and Demographic Characteristics of the Districts

			ELL-related variables					Ethnicity							
District ID	Enroll-ment size	Number of ELLs	% ELL	% ELL broad	% Exit 4	% Exit 5	% Exit 6	% Native Americans	% Asian or Pacific Islanders	% African Americ an	% Hispanic	% White	% Disability	% FRL	% Home-less
D042	4692	1491	31.8	44.1	4.2	4.9	6.3	1.4	3.0	18.1	58.3	19.2	15.6	67.2	1.3
D017	2283	1039	45.5	49.9	5.7	4.7	2.7	1.0	3.3	20.7	51.8	23.2	13.7	63.7	0.4
D004	2900	453	15.6	18.7	2.0	1.7	2.2	0.9	5.6	2.8	29.2	61.4	9.4	34.6	1.0
D080	6077	451	7.4	8.8	1.2	0.8	0.7	1.2	4.3	2.0	17.5	75.1	9.1	25.3	1.4
D170	1265	411	32.5	37.2	5.5	5.4	2.2	0.6	0.7	0.7	54.2	43.8	11.4	56.0	0.6
D014	3590	309	8.6	17.9	2.0	0.9	0.9	0.5	6.7	15.2	12.8	64.7	11.4	24.2	0.0
D009	688	275	40.0	45.3	4.7	1.9	2.8	1.0	7.1	2.8	64.0	25.1	11.6	67.6	5.4
D027	1600	267	16.7	19.3	1.4	1.4	5.1	0.8	3.9	1.8	26.8	66.8	7.0	30.5	1.9
D005	437	240	54.9	58.8	10.8	4.6	4.3	1.4	0.0	1.6	81.7	15.3	13.0	79.2	4.6
D092	1731	213	12.3	14.7	2.1	1.3	2.3	0.8	3.5	2.2	15.1	78.5	8.8	24.4	0.8
D028	1924	197	10.2	13.2	0.8	1.4	1.3	0.7	7.9	1.7	12.2	77.5	9.5	15.4	1.9
D006	869	179	20.6	25.0	1.7	2.0	1.6	1.0	4.1	2.5	42.1	50.2	9.0	39.9	2.6
D002	386	153	39.6	46.6	4.4	2.8	3.1	1.3	3.9	2.1	63.0	29.8	6.2	66.8	2.6
D055	2012	149	7.4	9.5	1.6	0.6	0.7	1.8	3.3	11.4	20.3	63.2	8.4	42.8	0.3
D045	362	142	39.2	40.9	5.2	2.2	5.5	0.8	0.0	0.0	50.8	48.3	13.3	34.3	0.6
D052	791	111	14.0	16.4	1.5	0.5	0.4	2.7	5.3	25.9	34.0	32.1	9.5	64.7	0.8
D069	335	103	30.7	41.8	5.4	5.1	2.4	0.3	1.2	0.0	44.5	54.0	9.3	34.3	0.0
D044	3789	101	2.7	4.2	0.3	0.3	0.3	0.6	4.2	2.6	8.7	83.9	7.7	7.3	0.1

District ID	Enroll-ment size	Number of ELLs	% ELL	% ELL broad	% Exit 4	% Exit 5	% Exit 6	% Native Americans	% Asian or Pacific Islanders	% African American	% Hispanic	% White	% Disability	% FRL	% Home-less
			ELL-related variables					Ethnicity							
D119	223	84	37.7	44.8	8.1	6.3	0.0	0.0	0.0	0.9	57.8	41.3	9.4	62.8	2.2
D172	162	81	50.0	60.5	6.2	5.6	6.8	1.2	1.9	0.6	66.0	30.2	4.3	66.7	0.0
D113	212	76	35.8	38.2	0.0	22.2	0.0	25.5	0.5	0.9	15.6	57.5	8.5	56.6	0.0
D070	297	73	24.6	33.7	2.4	3.4	4.4	1.0	0.0	0.3	41.4	57.2	7.7	44.4	0.0
D116	411	70	17.0	17.3	1.5	2.2	1.2	1.2	1.2	0.7	27.0	69.8	11.4	50.4	0.7
D093	1078	53	4.9	6.1	1.1	0.5	0.9	0.7	1.6	0.8	13.6	83.2	9.3	27.6	0.4
D110	1421	49	3.4	6.3	0.4	1.0	0.5	1.4	0.8	1.5	18.0	78.3	12.3	40.0	2.3
D013	108	44	40.7	46.3	13.0	5.6	2.8	2.8	6.5	8.3	60.2	22.2	13.0	81.5	0.9
D139	1195	44	3.7	9.2	0.4	0.0	0.3	1.0	0.7	3.2	60.0	35.1	10.9	72.4	7.5
D157	188	43	22.9	25.0	2.7	0.0	4.8	0.0	0.0	0.0	24.5	75.5	17.6	29.8	3.7
D181	212	42	19.8	23.1	2.4	1.9	2.4	1.9	4.7	5.7	36.3	51.4	5.7	31.1	0.0
D015	1118	35	3.1	3.9	0.7	0.0	0.6	0.6	2.7	1.9	11.4	83.5	11.4	13.3	0.4
D165	139	33	23.7	31.7	0.7	3.6	1.4	0.7	0.0	0.7	55.4	43.2	11.5	47.5	0.7
D167	155	33	21.3	27.1	2.6	1.9	3.9	1.3	0.6	0.6	34.8	62.6	10.3	49.7	0.0
D041	368	28	7.6	9.5	0.3	2.4	2.2	1.1	0.8	0.3	15.8	82.1	9.2	43.2	0.0
D088	74	28	37.8	39.2	10.8	2.7	5.4	0.0	0.0	0.0	60.8	39.2	8.1	52.7	0.0
D140	712	28	3.9	5.2	0.1	0.4	1.1	1.0	0.4	2.0	27.5	69.1	7.7	33.1	0.1
D012	217	27	12.4	13.4	0.9	0.9	0.5	1.8	3.7	4.6	29.0	60.8	12.0	38.7	2.3
D136	133	22	16.5	24.1	1.5	2.3	4.5	0.0	0.0	0.8	53.4	45.9	4.5	66.9	0.8
D063	852	21	2.5	4.0	0.2	0.1	0.6	1.6	3.4	9.3	17.3	68.4	8.5	17.6	0.0

We calculated reclassification rates, which is equal to the sum of reclassified student percentages in the three grades divided by the ELL percentage from Table 10. Figure 1 displays a scatter plot of reclassification rates against the number of ELLs. For most districts, the reclassification rates range between 20% and 50%. A greater range of reclassification rates is shown among districts with a smaller number of ELLs. As can be seen from Figure 1, a few districts with reclassification rates lower than 20% as well as a couple of districts with the rates higher than 50% have a very small number of ELL students in the cohort, which is less than 100-150 ELL students.

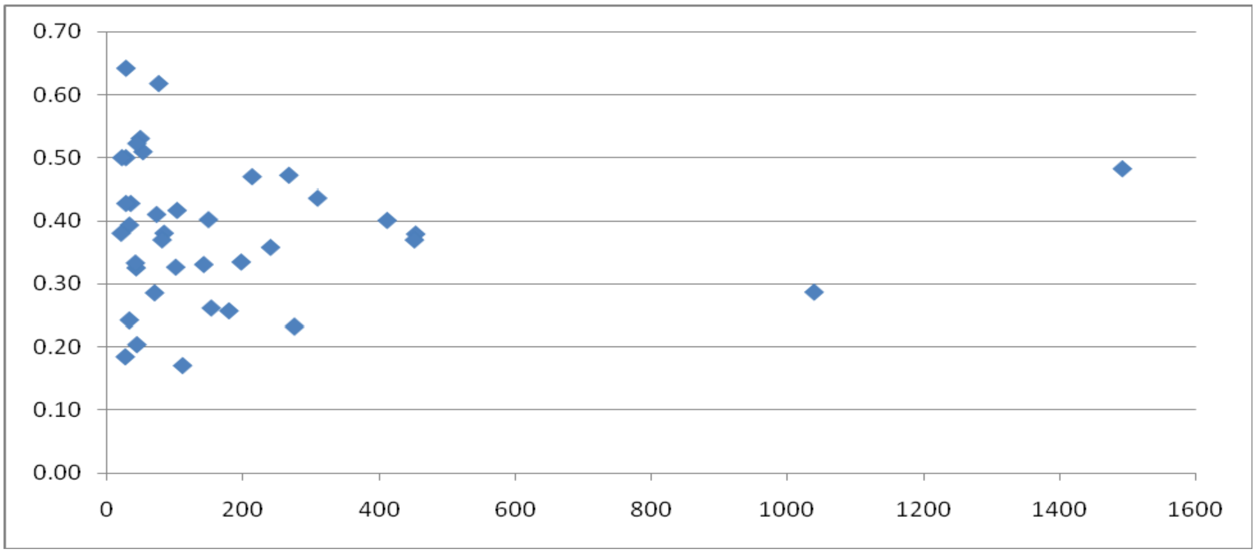

Figure 1. A scatter plot of reclassification rates against the number of ELLs.

Table 11 presents the descriptive statistics of students who are identified as Reclassified during Grades 4, 5, and 6. Reclassified ELL students' demographics are, in general, similar to those of other ELL students. Students who were reclassified during earlier grades consist of a slightly higher percentage of Asians and Pacific Islanders and a lower percentage of Hispanics when compared with students who are reclassified during later grades. Also, students who were reclassified during earlier grades have slightly lower percentages of economically disadvantaged students: 75% for the students who are reclassified during Grade 4; 78% for those reclassified during Grade 5; and 81% for those reclassified during Grade 6. The proportion of economically disadvantaged students for all ELL students is 81% for the study sample (see Table 11).

Table 11

Demographic Information for ELL Students Who Are Reclassified during Grades 4, 5, and 6

Demographics	Reclassified in Grade 4 (*n*=993)	Reclassified in Grade 5 (*n*=864)	Reclassified in Grade 6 (*n*=905)
Native American	0.0%	3.0%	0.0%
Asian/Pacific Islander	10.0%	8.0%	8.0%
Black	2.0%	1.0%	1.0%
Hispanic	81.0%	83.0%	88.0%
White	7.0%	6.0%	4.0%
Disability status	3.0%	4.0%	7.0%
Migrant status	2.0%	1.0%	2.0%
Immigrant status	1.0%	1.0%	1.0%
Economically disadvantaged	75.0%	78.0%	81.0%
Homeless status	1.0%	1.0%	1.0%

Methods

Study Design

This study compares post-reclassification growth of reclassified ELLs with growth of other students in the same period of time in reading and math in order to assess whether the given reclassification system and criteria was beneficial to ELL students. Design components (Shadish et al., 2002) we employ in this study to make the comparison more meaningful include repeated measures, multiple comparison groups, and multiple outcomes.

Repeated measures and the application of appropriate statistical methods (see the following section for growth modeling techniques) help track the same group of individuals over time, partitions variability in growth outcomes due to time-specific variations or measurement errors and to inter-individual variability, and opens up the possibility of testing key relationships of interest.

Use of multiple control comparison is an effective design feature that rules out many confounding issues that may arise in longitudinal studies. In the absence of control groups, growth patterns that we observe may be due to maturation, instrumentation, or history (see Table 2.4 in Shadish et al., 2002, p.55 for a summary of threats to internal validity), which may be erroneously interpreted as effects of treatment (e.g., reclassification) on time series in outcomes. For example, without comparison groups, naturally slower growth between Grade 7 and Grade 8 in reading due to maturation and/or characteristics of curricula materials (threat of maturation) may be erroneously considered as a negative effect of reclassification. Furthermore, there may be a slight upward bump between Grade 7 and Grade 8 due to the ways that vertical equating is conducted, which may also lead to misinterpretation (threat of instrumentation). This study uses the following multiple comparison groups: the entire non-ELL group and the other ELLs who are not reclassified during Grades 4, 5, or 6. The first comparison group, all non-ELLs in the cohort, can serve as reference for growth trajectories that show how diverse students grow over grades in reading and math achievement during the studied period. The second comparison group, the other ELLs who were not reclassified, might help give an idea of how ELL students would grow without reclassification during the studied period. As the comparison groups are nonequivalent groups to the target group (i.e., ELL students who are reclassified in specific grades), the use of multiple control groups is especially useful.

The study design is also complemented by the use of multiple outcomes. We use reading and math achievement scores in this study to serve as multiple outcomes. When a general pattern that is fundamental concerning inferences on ELL reclassification (e.g., more rapid growth rates of reclassified ELLs) is consistent across results between reading and math outcomes, we have more reason to claim that enhanced academic growth is due to reclassification. This means that the inference is further strengthened.

As noted earlier, although we take this view of considering the underlying reclassification system and criteria as an intervention, our study is limited in its internal validity and thus any intention to make causal inferences should be tempered. A key limitation of internal validity lies in the fact that comparison groups are nonequivalent groups. Since we compare groups that are different in many respects, we cannot be sure whether the study's findings are due to reclassification or preexisting differences in many characteristics across the groups. The use of an ITS design could have helped greatly with increasing the internal validity of the study. However, within the growth modeling framework, we could not implement the ITS design because of its limited feasibility in studies tracking state assessment data mainly for two reasons. First, although state assessment scores tend to have a monotonically increasing function over grades (see Kim & Herman, 2009; also see descriptive statewide means provided in state department of education websites), they do not necessarily follow certain functional forms such as linear or quadratic growth. Without functional forms, the application of ITS would not be possible or meaningful. Secondly, even when reasonably simple functional forms can be estimated, the number of repeated measures tend to be too short to apply ITS to the data in a meaningful way. For example, as will be seen, even when we had six-year annual assessment data from Grades 3 to 8, academic growth tends to be discontinuous without any intervention due to different rates of growth between earlier and later grades. In such cases, it was not feasible to estimate additional discontinuity due to an intervention.

The external validity of the current study is enhanced through careful selection of the sample and data. We use statewide individual-level student data from two states linked across several years. These data enable us to examine the entire state population, which ensures the external validity of our findings within the states. We also chose states with a range of reclassification criteria, which will permit studies of various types of reclassification criteria as currently adopted by states or local districts or schools.

Analysis Methods

We will use a growth modeling technique (see Diggle, Liang, & Zeger, 1994; Raudenbush & Bryk, 2002; Singer & Willet, 2003) to examine growth trajectories in academic achievement over grades. Growth modeling techniques have been widely applied in various fields, including education, medicine, and psychology. In the growth modeling framework, within-individual models estimate growth parameters for each individual and between-individual models allows for studies of individual differences in terms of growth parameters.

From this broad class of hierarchical modeling (HMs) or multilevel models, we will use model specifications that best suit our research questions and the data at hand, such as modeling discontinuous individual growth and latent variable regressions in a growth modeling framework. First, in reading growth, the models assume differential growth rates between elementary school grades (Grades 3-5) and middle school grades (Grades 5-8), since students tend to grow more rapidly during earlier grades than later grades. This entails piece-wise growth modeling (Raudenbush & Bryk, 2002, pp. 178-179) or modeling discontinuous individual growth (Singer & Willet, 2003, Chapter 6).

Secondly, it is important to hold constant the performance status prior to reclassification, while comparing post-reclassification/middle-school growth rates of different groups by ELL status, since they start at vastly different levels. When different groups start at appreciably different levels, it may be that they tend to grow at different rates. In such cases, it would not be meaningful to compare growth rates across the groups without taking into account their prior status. This involves latent variable regression in a growth modeling framework (Choi & Seltzer, 2010; Muthen & Curran, 1997; Seltzer, Choi, & Thum, 2003).

Specifically, Model 1 (see the following text) is used to analyze the entire sample to estimate growth trajectories of student groups by ELL status: students reclassified at Grades 4, 5, and 6, respectively; ELL students who are not reclassified in the above three grades (the "OtherELL" variable is the indicator variable); and non-ELL students (the "nonELL" variable is the indicator variable). Equation 1(a) is the within-individual model for reading, in which we model discontinuous growth rates between elementary and middle school grades by using two time-measuring variables. The “Elementary_Grade” variable is coded as a one time variable with values of -2, -1, and 0 respectively for Grades 3, 4, and 5. The “Middle_Grade” variable is coded as the other time variable with values of 0, 1, 2 and 3 respectively for Grades 5, 6, 7, and 8. With such a coding scheme, the intercept π_{0i} is the reading achievement status at Grade 5 for student *i*, the first slope π_{1i} is the growth rate during elementary school grades for student *i*, and the second slope is the growth rate during middle school grades for student *i*. In math, we track students from Grades 5 to 8; we use only one time variable, as can be seen in Equation 2(b). The intercept π_{0i} is the achievement status at Grade 5 for student *i* as it was in reading, the slope π_{1i} is the growth rate during middle school grades for student *i*.

Model 1 - Reading

$$Y_{ti} = \pi_{0i} + \pi_{1i}(\text{Elementary\_Grade})_{ti} + \pi_{2i}(\text{Middle\_Grade})_{ti} + e_{ti} \qquad 1(a)$$

$$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Exit4})_i + \beta_{02}(\text{Exit5})_i + \beta_{03}(\text{Exit6})_i + \beta_{04}(\text{OtherELL})_i + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Exit4})_i + \beta_{12}(\text{Exit5})_i + \beta_{13}(\text{Exit6})_i + \beta_{14}(\text{OtherELL})_i + r_{1i}$$

$$\pi_{2i} = \beta_{20} + \beta_{21}(\text{Exit4})_i + \beta_{22}(\text{Exit5})_i + \beta_{23}(\text{Exit6})_i + \beta_{24}(\text{OtherELL})_i + \beta_{25}(\beta_{00} - \pi_{0i}) + r_{2i} \qquad 1(b)$$

Model 1 - Math

$$Y_{ti} = \pi_{0i} + \pi_{1i}(\text{Middle\_Grade})_{ti} + e_{ti} \qquad 2(a)$$

$$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Exit4})_i + \beta_{02}(\text{Exit5})_i + \beta_{03}(\text{Exit6})_i + \beta_{04}(\text{OtherELL})_i + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Exit4})_i + \beta_{12}(\text{Exit5})_i + \beta_{13}(\text{Exit6})_i + \beta_{14}(\text{OtherELL})_i + \beta_{15}(\beta_{00} - \pi_{0i}) + r_{1i} \qquad 2(b)$$

The equations 1(b) and 2(b) in Model 1 are the between-individual model for reading and math respectively. The status at Grade 5 and growth rates are modeled as a function of binary indicators of ELL status groups, with the non-ELL group serving as a baseline. Thus, the parameters βs in the between-individual models 1(b) and 2(b) estimate differences in growth parameters between the non-ELL group and each of the other groups. Note that in modeling middle-school/post-reclassification growth rates, we use a modeling feature that allows for regressions among latent variables, as noted earlier. The middle-school/post-reclassification growth rate is regressed on achievement status at Grade 5 as well as on indicators of ELL status groups. In doing so, we can see the difference in student growth rates over post-reclassification/middle-school grades between reclassified ELL or ELL groups and non-ELL groups, holding constant their prior achievement status.

Model 2 and Model 3 are used to analyze the sample with only reclassified ELLs in order to see what factors are related to more rapid growth rates over post-reclassification grades. Both Model 2 and Model 3 share the same within-individual model as Model 1. Model 2 is concerned with student characteristics that may be associated with subsequent success. Thus, in Model 2, the between-individual model is specified as a function of binary indicators of the ethnicity categories found in state data (with White reclassified ELL students as the baseline group) and the economically disadvantaged. We also include the indicators for groups that are reclassified in different grades, with students reclassified at Grade 5 as the base group.

Model 2 - Reading

$$Y_{ti} = \pi_{0i} + \pi_{1i}(\text{Elementary\_Grade})_{ti} + \pi_{2i}(\text{Middle\_Grade})_{ti} + e_{ti} \quad 3(a)$$

$$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Native})_i + \beta_{02}(\text{Asian})_i + \beta_{03}(\text{Black})_i + \beta_{04}(\text{Hispanic})_i + \beta_{05}(\text{Exit4})_i + \beta_{06}(\text{Exit6})_i + \beta_{07}(\text{LowSES})_i + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Native})_i + \beta_{12}(\text{Asian})_i + \beta_{13}(\text{Black})_i + \beta_{14}(\text{Hispanic})_i + \beta_{15}(\text{Exit4})_i + \beta_{16}(\text{Exit6})_i + \beta_{17}(\text{LowSES})_i + r_{1i}$$

$$\pi_{2i} = \beta_{20} + \beta_{21}(\text{Native})_i + \beta_{22}(\text{Asian})_i + \beta_{23}(\text{Black})_i + \beta_{24}(\text{Hispanic})_i + \beta_{25}(\text{Exit4})_i + \beta_{26}(\text{Exit6})_i + \beta_{27}(\text{LowSES})_i + \beta_{28}(\beta_{00} - \pi_{0i}) + r_{2i} \quad 3(b)$$

Model 2 - Math

$$Y_{ti} = \pi_{0i} + \pi_{1i}(\text{Middle\_Grade})_{ti} + e_{ti} \quad 4(a)$$

$$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Native})_i + \beta_{02}(\text{Asian})_i + \beta_{03}(\text{Black})_i + \beta_{04}(\text{Hispanic})_i + \beta_{05}(\text{Exit4})_i + \beta_{06}(\text{Exit6})_i + \beta_{07}(\text{LowSES})_i + r_{0i}$$

$$\pi_{1i} = \beta_{10}\ \beta_{10} + \beta_{11}(\text{Native})_i + \beta_{12}(\text{Asian})_i + \beta_{13}(\text{Black})_i + \beta_{14}(\text{Hispanic})_i + \beta_{15}(\text{Exit4})_i + \beta_{16}(\text{Exit6})_i + \beta_{17}(\text{LowSES})_i + \beta_{18}(\beta_{00} - \pi_{0i}) + r_{1i} \quad 4(b)$$

Model 3 examines whether any districts are associated with more or less rapid post-reclassification growth rates. Thus, in Model 3, the between-individual model is specified as a function of binary indicators of 37 districts, with the largest district as the baseline district. We also included the indicators for groups that are reclassified in different grades, with students reclassified at Grade 5 as the base group.

Model 3 - Reading

$$Y_{ti} = \pi_{0i} + \pi_{1i}(\text{Elementary\_Grade})_{ti} + \pi_{2i}(\text{Middle\_Grade})_{ti} + e_{ti} \quad 5(a)$$

$$\pi_{0i} = \beta_{00} + \Sigma_{k=1 \text{ to } 37}\, \beta_{0k}(\text{DistI}_k)_i + \beta_{0,38}(\text{Exit4})_i + \beta_{0,39}(\text{Exit6})_i + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \Sigma_{k=1 \text{ to } 37}\, \beta_{1k}(\text{DistI}_k)_i + \beta_{1,38}(\text{Exit4})_i + \beta_{1,39}(\text{Exit6})_i + r_{1i}$$

$$\pi_{2i} = \beta_{20} + \Sigma_{k=1 \text{ to } 37}\, \beta_{2k}(\text{DistI}_k)_i + \beta_{2,38}(\text{Exit4})_i + \beta_{2,39}(\text{Exit6})_i + \beta_{2,40}(\beta_{00} - \pi_{0i}) + r_{2i} \quad 5(a)$$

Model 3 - Math

$$Y_{ti} = \pi_{0i} + \pi_{1i}(\text{Middle\_Grade})_{ti} + e_{ti} \quad 6(a)$$

$$\pi_{0i} = \beta_{00} + \Sigma_{k=1 \text{ to } 37}\, \beta_{0k}(\text{DistI}_k)_i + \beta_{0,38}(\text{Exit4})_i + \beta_{0,39}(\text{Exit6})_i + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \Sigma_{k=1 \text{ to } 37}\, \beta_{1k}(\text{DistI}_k)_i + \beta_{1,38}(\text{Exit4})_i + \beta_{1,39}(\text{Exit6})_i + \beta_{1,40}(\beta_{00} - \pi_{0i}) + r_{1i} \quad 6(a)$$

In specifying Model 2, we included variables that were known to be significantly related to performance status from previous literature, exploratory analysis, and findings from Model 1 (e.g., grades when students are reclassified). Instead of including binary indicators of all districts as in Model 3, an alternative specification of the model (i.e., a three-level multilevel model that adds a level of nesting clusters, districts, in a multilevel modeling framework) is used in order to examine district factors (see, Raudenbush & Bryk, 2002). This alternative model can be considered as an omnibus test to see whether there is no significant variability across districts, and take both results from this and Model 3 into consideration to draw our conclusion. We will discuss the results of this alternative specification, as well as the results of Model 3 in the following section.

Results

Expected Growth Trajectories of Reclassified ELLs as Compared to Non-ELLs or Other ELLs

Table 12 presents the results for achievement growth trajectories in reading. Regardless of ELL status, students tend to grow more rapidly in elementary school grades (i.e., Grades 3 to 5) than in middle school grades (i.e., Grades 5 to 8) in reading achievement. Thus, student discontinuous growth between elementary and middle school grades is captured by two distinct growth rate parameters as explained in Model 1. In what follows, we interpret parameters in the first panel of Table 12 (the panel of Fixed Effects) and compare the estimated average growth trajectories among different ELL groups.

Table 12

Results from Model 1-Reading

Fixed effects	Coefficient	*SE*	Ratio	*p*-value
Model for status at Grade 5				
Intercept	622.12	0.33	1883.4	0.000
EXIT4	-16.47	2.04	-8.1	0.000
EXIT5	-29.85	2.17	-13.7	0.000
EXIT6	-45.94	2.12	-21.6	0.000
OTHEREL	-89.92	1.04	-86.6	0.000
Model for growth rate during Elementary (Grades 3-5)				
Intercept	26.12	0.12	211.7	0.000
EXIT4	2.92	0.76	3.9	0.000
EXIT5	5.02	0.80	6.2	0.000
EXIT6	4.25	0.79	5.4	0.000
OTHEREL	-0.35	0.40	-0.9	0.391
Model for growth rate during Middle, Post-reclassification (Grades 6-8)				
Intercept	10.65	0.07	153.0	0.000
EXIT4	1.05	0.39	2.7	0.007
EXIT5	0.92	0.41	2.2	0.025
EXIT6	2.08	0.40	5.2	0.000
OTHEREL	2.93	0.23	12.5	0.000
Status at Grade 5	-0.56	0.01	-47.8	0.000
Random effects	**Variance**	***SE***	**Ratio**	***p*-value**
Level-1 variance, temporal (within student)				
Grade 3	1532.80	17.30	88.6	0.000
Grade 4	688.00	6.90	99.1	0.000
Grade 5	633.30	7.90	80.5	0.000
Grade 6	822.70	7.10	116.0	0.000
Grade 7	584.60	5.60	104.3	0.000
Grade 8	364.30	7.00	52.0	0.000
Level-2 variance (between student)				
Status at Grade 5	3557.40	27.80	127.8	0.000
Growth rate Elementary	89.60	5.00	17.8	0.000
Growth rate Middle	50.30	1.30	38.7	0.000

During elementary school grades (Grades 3 to 5), all reclassified ELL students tend to grow more rapidly than non-ELL students. ELL students reclassified at Grade 4 grow more rapidly on average by 2.9 points annually; ELL students reclassified at Grade 5 grow more rapidly on average by 5.0 points annually; ELL students reclassified at Grade 6 grow more rapidly on average by 4.3 points annually. However, other ELL students do not grow faster than non-ELL students (coeff = -0.35, *p*-value = 0.40).

At Grade 5, there are significant gaps in achievement between non-ELL students and ELL students. Compared to non-ELL students, students reclassified at Grade 4 show an average estimated gap of 16 points; students reclassified at Grade 5 show an average gap of 30 points; students reclassified at Grade 6 show an average gap of 46 points. Other ELL students show an average gap of 90 points when compared to non-ELL students. From the estimated inter-individual variability of status at Grade 5, one *SD* is approximately 60 points (the square root of 3557.40). Therefore, the estimated achievement gaps are about 0.3 *SD*s, 0.5 *SD*s, 0.8 *SD*s, and 1.5 *SD*s, respectively for ELL students reclassified at Grade 5, Grade 6, Grade 7, and ELL students who are not reclassified. All of these are significant differences, with the sizes ranging from moderate to very large differences. One can also see a fair amount of differences in achievement at Grade 5 across students who are reclassified in different grades. Students who are reclassified in later grades show significantly lower levels of achievement than students reclassified in earlier grades.

During the middle school grades (Grades 6 to 8), even after controlling for status at Grade 5, all ELL student groups tend to grow more rapidly than non-ELL students. ELL students reclassified at Grade 4 grow more rapidly on average by 1.1 points annually; ELL students reclassified at Grade 5 grow more rapidly on average by 0.9 annually; ELL students reclassified at Grade 6 grow more rapidly on average by 2.1 annually. It is notable that, unlike the elementary school grades, other ELL students also show significantly more rapid growth rate than non-ELLs (by 2.9 annually).

Other ELLs tend to start very low in Grade 3, and do not show any tendency to catch up to non-ELL students during elementary school grades. However, they show significantly more rapid growth rates during middle school grades than non-ELL students holding constant achievement status at Grade 5. This should not be understood as other ELL students doing better than non-ELL students without further investigation, since their more rapid growth rates may be an artifact arising from various issues. For example, their more rapid growth rates may be an artifact of vertical scaling of scores. Since scores near the minimum scores seem to compress upward more over grades (see discussion around Tables 2 and 3) and since this specific group (other ELLs) has significantly more proportion of students with minimum or close to minimum scores, the estimated average growth rate must have been inflated. In addition, since this group includes more proportion of students who started at a very low level, there may be more regression toward mean over time for this group than for others.

To summarize, the reclassified ELL students tend to grow more rapidly than non-ELL students in the elementary school grades. Although they show a trend of catching up to their non-ELL peers, they tend to finish the elementary grades with significant magnitudes of achievement gaps, with the magnitudes being different for students who were reclassified in different grades. During the middle school grades, even after controlling for the achievement status at Grade 5, reclassified ELL students still tend to show more rapid growth rates than their non-ELL peers. This implies that ELLs reclassified at Grades 4, 5, or 6 tend to be the children who catch up with their non-ELL peers before reclassification, exiting with a certain amount of achievement gaps, but still continue to catch up to their non-ELL peers after reclassification.

Figure 2 displays the estimated average growth trajectories for different ELL groups. Figure 2 additionally displays four achievement level categories for all students, as designated by the state, which are Unsatisfactory, Partially Proficient, Proficient, and Advanced. The upper band of the figure shows the Proficient category, while the lower band shows the Partially Proficient category. Any scores higher than the upper band are in the Advanced category; likewise, any scores lower than the lower band are in the Unsatisfactory category. Since the scale scores are vertically equated across grades in this state, the achievement category bands move up as the grades go up.

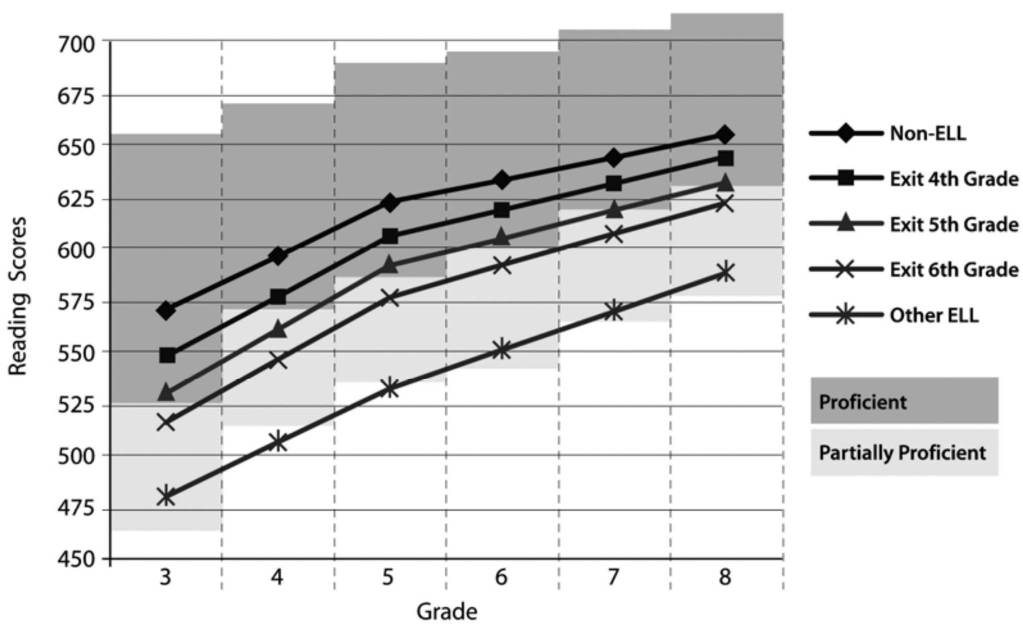

Figure 2. Estimated reading growth trajectories by ELL status and reclassified ELL status.

From this figure that shows the estimated trajectories superimposed on the achievement-level bands, one can see that, on average, there was no dramatic catch up of reclassified ELLs with their non-ELLs, although reclassified ELLs tend to show significantly higher growth rates than non-ELLs. For example, students who exited at Grade 5 barely achieved the Proficient category in Grade 3, and in Grade 4, these students achieved below the cut-off for the Proficient category and were in the Partially Proficient category. These students on average still barely achieved proficiency from Grades 6 to 8 after their reclassification.

Therefore, we see a discrepancy in findings when we judge trajectories from the estimated difference in growth rates and when the estimated trajectories are compared to state-designated achievement categories. This may complicate the interpretation of the results. Figure 2 gives some absolute sense of the achievement levels and trends of reclassified ELLs, based on state-designated achievement categories. Therefore, this may suggest that that, although their growth rates during the post-reclassification period was significantly more rapid relative to those of non-ELL students during the same period, the magnitude of differences may not be large enough to actually catch up to non-ELL peers. On average, ELL students continued to grow so that they kept up their academic performance after reclassification.

Table 13 presents the results for achievement growth trajectories in mathematics. At Grade 5, there are significant gaps in achievement between non-ELL students and ELL students. Students reclassified at Grade 4 show the estimated average achievement gap of 19 points; students reclassified at Grade 5 show the average gap of 30 points; students reclassified at Grade 6 show the average gap of 43 points; and other ELLs show the estimated average gap of 83 points, as compared to non-ELL students. From the estimated inter-individual variability of status at Grade 5, one *SD* is approximately 70 points (the square root of 4936.10). Therefore, the estimated achievement gaps are about 0.3 *SD*s, 0.4 *SD*s, 0.6 *SD*s, and 1.2 *SD*s, respectively for ELL students reclassified at Grades 4, 5, 6, and other ELLs. All of these are significant differences with the sizes ranging from small to large magnitudes. However, in comparison to the achievement gaps in reading, these are relatively smaller magnitudes. In addition, a fair amount of differences in achievement are shown for students reclassified in different grades. These differences follow the same pattern in reading: students reclassified later show significantly lower levels of achievement than students reclassified at earlier grades.

Table 13

Results from Model 1-Math

Fixed effects	Coefficient	*SE*	Ratio	*p*-value
Model for status at Grade 5				
Intercept	528.18	0.40	1335.5	0.000
EXIT4	-18.61	2.42	-7.7	0.000
EXIT5	-29.96	2.55	-11.8	0.000
EXIT6	-42.70	2.48	-17.2	0.000
OTHEREL	-82.83	1.26	-65.6	0.000
Model for Growth rate during middle, post-reclassification (Grades 6-8)				
Intercept				
EXIT4	1.07	0.39	2.7	0.007
EXIT5	1.02	0.41	2.5	0.014
EXIT6	1.05	0.40	2.6	0.009
OTHEREL	1.48	0.23	6.5	0.000
Status at Grade 5	-0.68	0.01	0.0	0.000
Random effects	**Variance**	***SE***	**Ratio**	***p*-value**
Level-1 variance, temporal (within student)				
Grade 5	742.60	11.10	66.7	0.000
Grade 6	674.70	7.00	96.8	0.000
Grade 7	662.40	6.60	99.7	0.000
Grade 8	259.60	7.90	33.0	0.000
Level-2 variance (between student)				
Status at Grade 5	4936.10	38.20	129.3	0.000
Growth rate Middle	59.10	1.60	37.4	0.000

During the middle school grades, even after controlling for the status at Grade 5, all ELL student groups tend to grow more rapidly than non-ELL students. ELL students reclassified at Grades 4, 5, 6 tend to grow more rapidly by 1 point, on average. Annually, other ELLs tend to grow 1.5 points more rapidly than non-ELLs; however, as noted in reading, this may be an artifact of vertical scaling over grades and should not be understood as other ELL students doing better than non-ELL students.

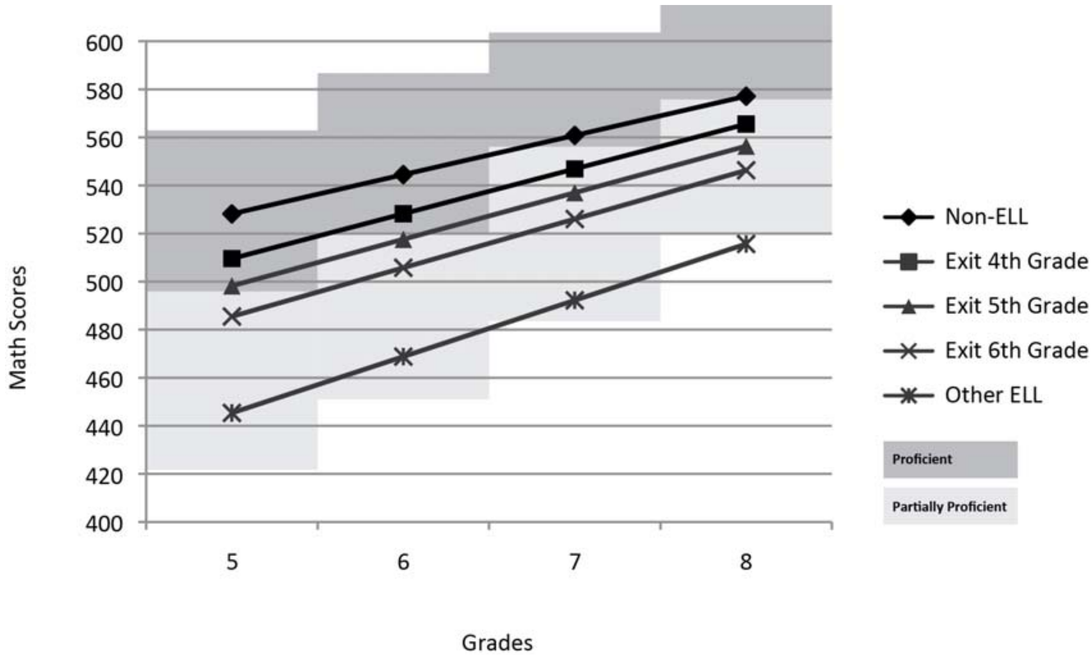

Figure 3. Estimated math growth trajectories by ELL status and reclassified ELL status
Note: The upper band of the shows the Proficient category, while the lower band shows the Partially Proficient category. Any scores higher than the upper band is the Advanced category; likewise, any scores lower than the lower band is the Unsatisfactory category. Since the scale scores are vertically equated across grades in this state, the achievement category bands move up as the grades go up.

Similar to the previous figure in reading (Figure 2), Figure 3 displays the estimated average growth trajectories for different ELL groups, overlaid on four achievement level categories for all students, as designated by the state, which are Unsatisfactory, Partially Proficient, Proficient, and Advanced.

The same finding emerges from math as from reading. From the Figure 3, one can see that there was no dramatic catch-up of reclassified ELLs to non-ELLs on average, although reclassified ELLs tend to show significantly higher growth rates than non-ELLs. For example, students who exited at Grade 5 barely achieved the Proficient category in Grade 3, and in Grade 4, these students achieved below the cut-off and were in the Partially Proficient category. These students on average still barely achieved at the Proficient category from Grades 6 to 8 after their reclassification.

Individual Differences in Estimated Growth Trajectories in Reading and Math

Results discussed so far were concerned with average growth trajectories of ELLs and other student groups. While average estimates provide answers to important questions, it is important to note that students within the different groups are not all identical students. The bottom panels of Table 12 and Table 13 show the variability of growth parameters (i.e., growth rates in elementary schools, status at Grade 5, and growth rates in middle schools in reading, and the last two in math) across individual students.

In terms of reading achievement status at Grade 5, the estimated differences in average achievement at the end of Grade 5 between non-ELL students and other groups (i.e., reclassified ELLs, exiting at Grades 4, 5, and 6 respectively; and other ELLs) range from 16 points to 90 points. However, one *SD* of the inter-individual variability is about 60 points (the square root of 3557.40), and therefore, across individual students, the scores can range from -120 to 120 points from their estimated averages. As for math achievement status at Grade 5, the estimated differences in average achievement at the end of Grade 5 between non-ELL students and other groups range from 19 points to 83 points. However, one *SD* of the inter-individual variability is about 70 points (the square root of 4936.10), and therefore, across individual students, the scores can range from -140 to 140 points from their estimated averages.

Similar observations apply to growth rates. In reading, the estimated differences in growth rates in middle school grades between the non-ELL group and the other groups range from 1 to 3 points, after controlling for the status at Grade 5. However, one *SD* of the inter-individual variability is about 7 points, which means that the growth rates can range from -14 to 14 from their estimated averages. In math, the estimated differences in growth rates in middle school grades between the non-ELL group and the other groups range from 1 to 1.5 points, after controlling for the status at Grade 5. However, one *SD* of the inter-individual variability is about 7.7 points, which means that the growth rates can range from -15.4 to 15.4 from their estimated averages. These results imply that growth trajectories within groups can be far more diverse than the differences across groups both in reading and math.

Characteristics of Reclassified ELLs Associated with Greater Subsequent Academic Success in Reading and Math

The analysis has thus far compared reclassified ELLs as a group to other groups (i.e., non-ELL students or other ELLs). This section deals with analyses of reclassified ELL students only and thus attempts to highlight the within-group heterogeneity among reclassified ELLs. Before running Model 2, in order to quantify the within-group heterogeneity, we first ran an unconditional model in which we estimated growth parameters only and their variances and covariance.

In the unconditional model, there is significant variability in middle school growth rates, as well as two other growth parameters. The mean of the middle school growth rates for reclassified ELLs is 13.71, and the *SD* is 7.66.

Given such heterogeneity, the analyses for this section search for variables that would account for the variability among reclassified ELL students in middle-school or post-reclassification growth rates. In line with previous analyses, we first included achievement status before middle school started (i.e., status at Grade 5). This explained much of the variability in middle school growth rates, as much as 28.30% of the variability. After controlling for the status at Grade 5, the *SD* of the growth rates was reduced to 6.48 from 7.66 in the unconditional model. Although significant variability is accounted for by the status before middle school grades, appreciable variability among these students still remains. For example, the annual difference of 6.48 cumulates to about 20 points over three middle school grades. Reclassified ELL students with a growth rate of one *SD* above would surpass the trajectory of non-ELLs and outperform them during middle school grades. Reclassified ELL students with a growth rate of one *SD* below would fall way behind and might achieve as low as or lower than other ELLs during middle school grades. The above results were from math growth, and a very similar pattern of results was also found in reading.

Table 14 presents the results of Model 2 – Reading, which attempts to explain the inter-individual variability in growth parameters. In terms of level at Grade 5, when compared to white reclassified ELL students, reclassified ELL students in other ethnicities tend to perform significantly lower in reading at Grade 5, except for Asian reclassified ELL students. Black reclassified ELL students on average perform lower by 17.3 points; Hispanic reclassified students perform lower by 17.9 points; and Native American reclassified students perform lower by 29 points. Consistent with the previous findings (see Table 12), students reclassified at Grade 4 on average perform significantly better than students reclassified at Grade 5 (by 12 points), while students reclassified at Grade 6 perform significantly lower than those who were reclassified at Grade 5 (by 16 points). Students receiving free or reduced lunch perform significantly lower by 15 points.

Table 14

Results from Model 2 - Reading

Fixed effects	Coefficient	*SE*	Ratio	*p*-value
Model for status at Grade 5				
Intercept	618.85	4.46	138.9	0.000
NATIVE	-29.31	10.66	-2.8	0.006
ASIAN	4.93	5.23	0.9	0.346
BLACK	-17.30	9.50	-1.8	0.069
HISPANIC	-17.87	4.35	-4.1	0.000
EXIT4	11.92	2.36	5.0	0.000
EXIT6	-15.75	2.41	-6.5	0.000
LOWSES	-14.49	2.44	-5.9	0.000
Model for growth rate during elementary (Grades 3-5)				
Intercept	31.44	1.95	16.1	0.000
NATIVE	-2.92	4.66	-0.6	0.531
ASIAN	7.41	2.30	3.2	0.001
BLACK	-0.69	4.13	-0.2	0.867
HISPANIC	-0.04	1.91	0.0	0.984
EXIT4	-2.40	1.03	-2.3	0.020
EXIT6	-0.85	1.05	-0.8	0.414
LOWSES	-1.06	1.06	-1.0	0.315
Model for growth rate during middle, post-reclassification (Grades 6-8)				
Intercept	13.23	1.06	12.4	0.000
NATIVE	-1.38	2.34	-0.6	0.555
ASIAN	-0.59	1.08	-0.5	0.586
BLACK	-0.10	1.98	0.0	0.962
HISPANIC	-2.35	0.91	-2.6	0.010
EXIT4	0.48	0.49	1.0	0.329
EXIT6	0.79	0.50	1.6	0.111
LOWSES	-0.27	0.50	-0.5	0.597
Status at Grade 5	-0.85	0.05	-17.0	0.000

Random effects	Variance	*SE*	Ratio	*p*-value
Level-1 variance, Temporal (within student)				
Grade 3	1178.10	52.60	22.4	0.000
Grade 4	515.50	20.50	25.1	0.000
Grade 5	533.10	25.50	20.9	0.000
Grade 6	645.60	21.70	29.8	0.000
Grade 7	525.90	18.90	27.8	0.000
Grade 8	265.50	21.80	12.2	0.000
Level-2 variance (between student)				
Status at Grade 5	2170.60	69.20	31.4	0.000
Growth rate Elementary	114.10	16.00	7.1	0.000
Growth rate Middle	41.50	3.90	10.7	0.000

As for elementary school growth rates, students reclassified at Grade 4 tended to grow slower than students reclassified at Grade 5. Annually, Asian reclassified students tend to grow 7.4 points more rapidly annually than White reclassified students.

Also, there was one significant correlate of middle school growth rates after controlling for status at Grade 5. Hispanic reclassified students tend to grow 2.4 points slower than White reclassified students. Since other ethnic categories do not turn out to grow at significantly different rates than White reclassified students, this implies that Hispanic reclassified students on average grow significantly slower than other reclassified students. Hispanic students tend to grow about 10.8 points annually as compared to the baseline group of students (White) who grow 13.2 points annually.

Table 15 presents math results from Model 2 – Math, in which the same set of predictors for reading – ethnicity, grade levels at reclassification, and free or reduced lunch status – are included as predictors of growth parameters in math. Similar to the findings in reading, significant predictors emerged as expected in terms of status, but only the ethnicity category turned out as a significant predictor of growth rates.

Table 15

Results from Model 2 - Math

Fixed effects	Coefficient	*SE*	Ratio	*p*-value
Model for status at Grade 5				
Intercept	528.31	5.44	97.0	0.000
NATIVE	-49.37	13.04	-3.8	0.000
ASIAN	16.92	6.40	2.6	0.008
BLACK	-18.21	11.61	-1.6	0.117
HISPANIC	-24.64	5.33	-4.6	0.000
EXIT4	9.21	2.88	32.0	0.010
EXIT6	-12.45	2.91	-42.8	0.000
LOWSES	-11.80	2.96	-39.8	0.000
Model for growth rate during middle, post-reclassification (Grades 5-8)				
Intercept	18.84	1.14	16.5	0.000
NATIVE	-9.60	2.57	-3.7	0.000
ASIAN	0.96	1.18	0.8	0.418
BLACK	2.05	2.16	0.9	0.343
HISPANIC	-1.69	0.99	-1.7	0.089
EXIT4	-0.29	0.53	-0.5	0.586
EXIT6	-0.21	0.54	-0.4	0.692
LOWSES	-0.05	0.55	-0.1	0.934
Status at Grade 5	-0.67	0.05	-14.6	0.000
Random effects	**Variance**	***SE***	**Ratio**	***p*-value**
Level-1 variance, Temporal (within student)				
Grade 5	609.30	35.70	1704.8	0.000
Grade 6	589.30	23.60	2501.9	0.000
Grade 7	645.10	24.50	2634.2	0.000
Grade 8	287.90	28.20	1019.3	0.000
Level-2 variance (between student)				
Status at Grade 5	3250.90	100.40	32.4	0.000
Growth rate Middle	45.70	5.30	8.7	0.000

As compared to White reclassified ELL students, Black reclassified ELL students did not show significant difference in achievement status at Grade 5; Native American and Hispanic students tended to perform significantly lower in math achievement at Grade 5 by 49 points and 25 points, respectively; and the Asian reclassified ELL students, on average, performed significantly higher by 17 points. As noted in the previous findings (see Table 13), students reclassified at Grade 4 performed significantly better than students reclassified at Grade 5 (by 9 points), while students reclassified at Grade 6 performed significantly lower than those who were reclassified at Grade 5 (by 12 points). Students receiving free or reduced lunch performed significantly lower by 12 points.

Significant correlates of middle school growth rates for math after controlling for status at Grade 5 were also student ethnicity groups, as was found in reading. Hispanic reclassified students tend to grow 1.7 points slower annually than White reclassified students, which approaches significance (*p*-value = 0.09). Native American reclassified students tend to grow more slowly, by 9.6 points annually.

District Membership Associated with Greater Subsequent Academic Success of Reclassified ELLs in Reading and Math

Table 16 presents part of the results of Model 3 which examines ELL student growth by district membership. For brevity of presentation, we included only the estimates with respect to the middle school growth rates. As growth rates during the middle school years roughly correspond to growth rates during grades after reclassification, which is the primary outcome of the study, parameters related to middle school growth rates are of key interest.

Table 16

Excerpts from Results from Model 3 – Reading and from Model 3 – Math

	Reading			Math		
Fixed effects	Estimate	*SE*	*p*-value	Estimate	*SE*	*p*-value
D2	-1.07	(1.63)	0.51	-2.87	(1.73)	0.10
D4	1.08	(0.90)	0.23	-0.85	(0.95)	0.37
D5	2.01	(1.22)	0.10	4.52	(1.29)	0.00
D6	-1.89	(1.53)	0.22	-5.47	(1.62)	0.00
D9	-1.08	(1.36)	0.43	-2.24	(1.45)	0.12
D12	2.68	(4.81)	0.58	1.28	(5.13)	0.80
D13	0.93	(2.27)	0.68	-3.87	(2.43)	0.11
D14	1.33	(0.99)	0.18	2.34	(1.06)	0.03
D15	0.08	(2.82)	0.98	0.63	(3.00)	0.83
D17	0.72	(0.73)	0.32	4.17	(0.77)	0.00

	Reading			Math		
Fixed effects	Estimate	*SE*	*p*-value	Estimate	*SE*	*p*-value
D27	-0.48	(0.98)	0.63	-0.82	(1.04)	0.43
D28	1.08	(1.33)	0.42	3.10	(1.40)	0.03
D41	0.49	(2.41)	0.84	5.37	(2.54)	0.03
D44	-2.61	(1.77)	0.14	1.88	(1.89)	0.32
D45	5.27	(1.53)	0.00	3.87	(1.62)	0.02
D52	-1.04	(2.31)	0.65	-0.45	(2.42)	0.85
D55	-0.05	(1.41)	0.97	3.36	(1.50)	0.03
D63	-3.12	(3.61)	0.39	7.37	(3.84)	0.06
D69	1.53	(1.59)	0.34	1.19	(1.70)	0.49
D70	-0.65	(1.92)	0.73	2.05	(2.05)	0.32
D80	0.25	(0.90)	0.78	3.23	(0.95)	0.00
D88	5.85	(2.81)	0.04	8.71	(3.00)	0.00
D92	-0.43	(1.14)	0.71	-1.32	(1.21)	0.28
D93	0.84	(1.99)	0.67	2.01	(2.09)	0.34
D110	-1.00	(2.02)	0.62	2.24	(2.16)	0.30
D113	1.81	(1.67)	0.28	-3.60	(1.78)	0.04
D116	-1.85	(2.21)	0.40	3.40	(2.35)	0.15
D119	0.30	(1.89)	0.87	0.51	(2.01)	0.80
D136	2.70	(3.41)	0.43	8.02	(3.65)	0.03
D139	-7.88	(3.58)	0.03	-10.29	(3.82)	0.01
D140	5.02	(3.27)	0.12	4.51	(3.50)	0.20
D157	-6.43	(3.02)	0.03	-7.38	(3.22)	0.02
D165	7.11	(3.46)	0.04	4.48	(3.69)	0.23
D167	3.85	(2.83)	0.17	5.82	(3.02)	0.05
D170	-1.13	(0.89)	0.20	-3.90	(0.95)	0.00
D172	1.25	(1.93)	0.52	1.13	(2.06)	0.58
D181	-1.08	(2.69)	0.69	-2.07	(2.87)	0.47
EXIT4	0.51	(0.50)	0.30	-0.29	(0.53)	0.58
EXIT6	0.84	(0.51)	0.10	-0.06	(0.54)	0.91

In reading results, shown in the left panel of Table 16, there are only five districts among the 37 districts that show significantly different post-reclassification growth rates when compared to the base district which is the largest district. Among these five, three districts show positive differences, that is, more rapid growth rates, while two districts show negative differences, that is, slower growth rates. The three districts, D45, D88, and D165 show the estimated differences in post-reclassification annual growth rates of 5.3, 5.9 and 7.1 points, respectively. Two districts, D139 and D159, show the estimated differences of -7.9 and -6.4 points, respectively. One *SD* of variability across individuals was estimated to be 6.3 (square root of 39.2) in this model. Thus, most differences in both directions are around one *SD*, which is a fairly appreciable magnitude.

As for math results, shown in the right panel of Table 16, seventeen districts among the 37 districts show significantly different post-reclassification math growth rates when compared to the base district. Among these seventeen districts, twelve districts show positive differences, that is, more rapid growth rates, while five districts show negative differences, that is, slower growth rates. The twelve districts that show more rapid growth rates are as follows: D5 (4.5), D14 (2.3), D17 (4.2), D28 (3.1), D41 (5.4), D45 (3.9), D55 (3.4), D63 (7.4), D80 (3.2), D88 (8.8), D136 (8.0) and D167 (5.8). The estimated difference in post-reclassification annual growth rates are shown in the parentheses. One *SD* of variability across individuals was estimated to be 6.4 (square root of 41.5) in this model. Thus, most of the significant differences in both directions are from about less than half of one *SD* of inter-individual variability to more than one *SD*. Smaller differences in magnitudes turned out significant in math, which relates to the fact that more districts showed significant differences.

Not surprisingly, the districts that showed significant differences in reading are a complete subset of the districts that showed significance differences in math. Given that we may be prone to inflated Type I errors with testing many variables (i.e., 37 district indicators), we focus on the districts that emerged significant both in reading and math. However, these districts tend to be the ones with a relatively small number of reclassified ELLs (see Table 10), and with numbers that are too small, it is difficult to infer that the estimates are representative. With an exception of one district, D48 with forty-seven reclassified ELL students, all districts that showed significance had less than 15 reclassified ELLs, ranging from 8 to 14 ELLs.

Math results reveal that some of the districts that showed significant positive differences in middle school growth rates are districts with a large number of reclassified ELLs. These districts include D17 with 298 reclassified ELLs, D80 with 167 reclassified ELLs, and D14 with 135 reclassified ELLs, but these significances may be subject to inflated Type I errors.

We have alternatively fitted three-level multilevel models in which repeated measures are nested within individual students, who are in turn nested within districts. These may serve as omnibus tests to show that there is no significant variability across districts. The results suggest that there was no variability across districts in post-reclassification growth rates. Caution is needed in interpreting the results since the number of students in the districts are very different (see Table 10, columns Number of ELLs, % of Exit 4, % of Exit5, and % of Exit6). Most of the reclassified students in our sample come from a few districts, and the other districts include a very small number of the reclassified students. Multilevel models using districts as another cluster level may not be useful in capturing differences across districts in such settings, because the estimates for districts can tend to be pulled much toward the average which is mostly determined by districts that are very large (for more details on empirical Bayes estimators or shrinkage estimators, see Raudenbush and Bryk, 2002, Chapter 4).

Combining both results is expected to compensate for potential caveats of each other and have yielded consistent results. From the finding of Model 3, the districts that had significantly more rapid or slower post-reclassification growth rates in both subjects were mostly districts with very small numbers of reclassified ELLs, which also suggests that the district factor may not be contributing much in predicting post-reclassification growth rates.

Figures 4 and 5 display scatterplots of the estimated district middle school growth rates shown in Table 16 against the estimated district status at Grade 5 for reading and math respectively. The figures show district average estimates of where reclassified ELL students started at Grade 5, as well as district average estimates of how rapidly they grow. They additionally show the number of reclassified students, which are reflected by the size of bubbles. The base district is shown with the largest bubble (719 reclassifed ELL students). As can be seen from the figures, the base district has a lower range of status at Grade 5 and a middle range of middle school growth rates, relative to the other districts. Among the districts with significantly higher rapid growth rates both in reading and math, districts D88 and D165 are the ones with similar status at Grade 5. District D5 has an appreciably higher status at Grade 5 than the base district, and district D45 has one of the highest achievement statuses at Grade 5 among all districts. As for the three large districts with significantly more rapid growth rates in math only, all three districts, D17, D80, and D14, tend to have an appreciably higher prior status as compared to the base districts.

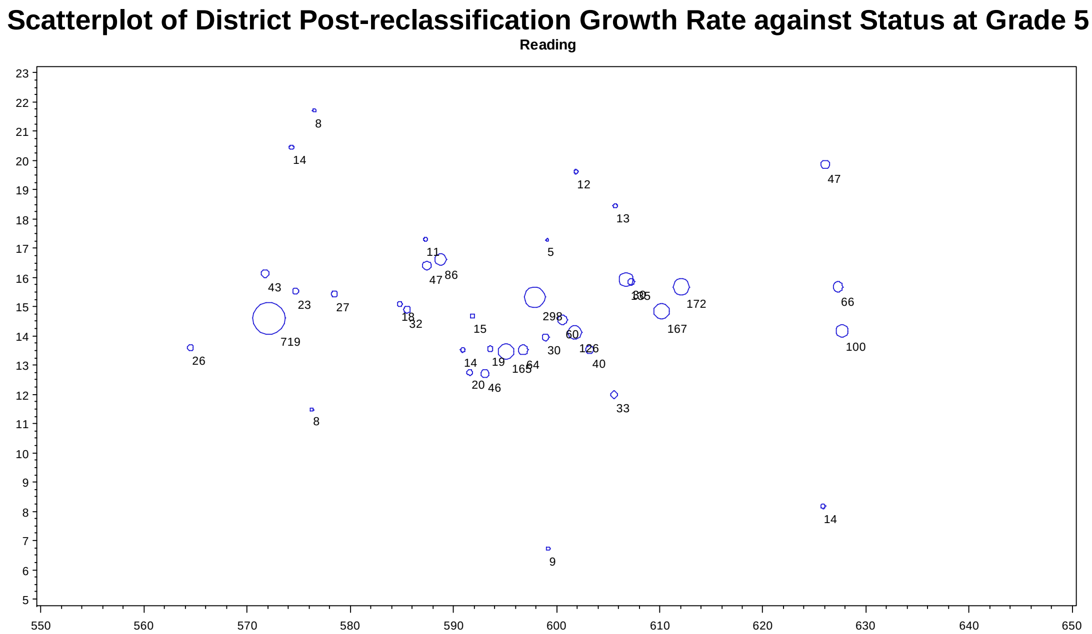

Figure 4. Scatterplot of district post-reclassification growth rate against status at Grade 5 in reading.

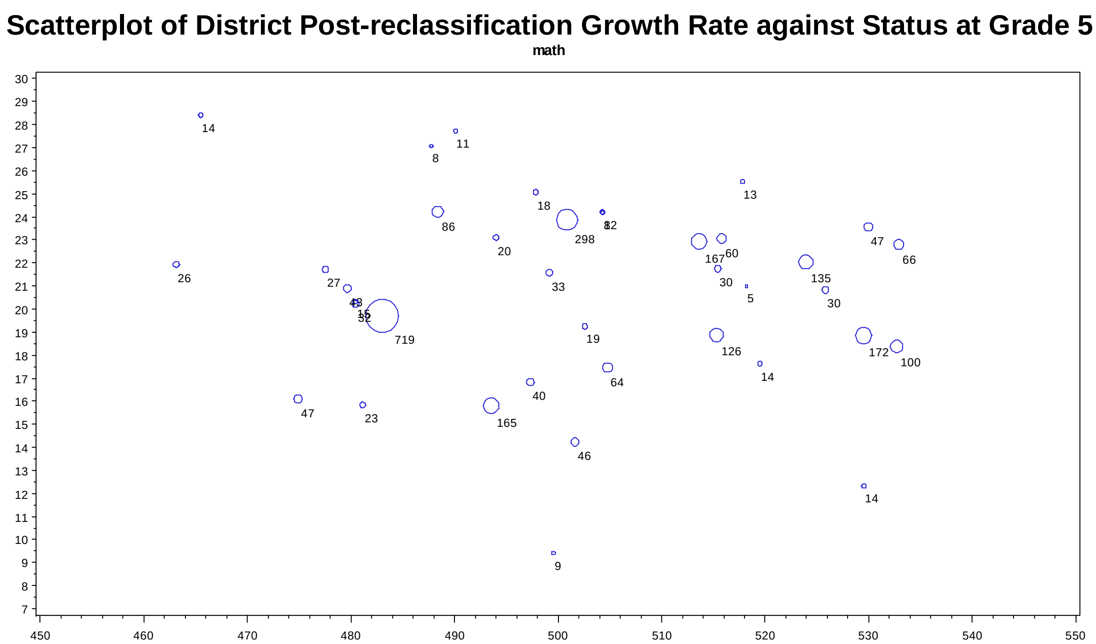

Figure 5. Scatterplot of district post-reclassification growth rate against status at Grade 5 in math.

Relationships of Levels of ELP to Subsequent Academic Success in Reading and Math

This section examines the relationships between ELP levels upon exiting and subsequent academic growth. The ELP data available to us contain far more complete data on ELP assessments in the more recent years but not for the data that went back several years. Due to the fact that we would not want the outcome of interest and the predictor to occur concurrently, we need the ELP level upon exiting or prior to exiting in order to investigate its relationship to a growth rate after reclassification.

Encountering this issue, we chose to limit our sample, instead of having to deal with too much missing data. For students reclassified at Grades 4 or 5, the data on the ELP levels upon exiting went farther back in comparison to students reclassified at Grade 6, two years and one year respectively, which had a vast amount of missing data. Thus, for this set of analyses, we focused on students who were reclassified at Grade 6. For these students, 6.5% had missing data in either ELP levels or reading proficiency levels upon exiting and this amount of missing data was acceptable. In our analyses, we assumed that the 6.5 % is missing at random.

Table 17

Crosstabs of ELP Levels Against Reading Proficiency Levels

	ELP proficiency levels					
Reading proficiency levels	1	2	3	4	5	Total
1 (Unsatisfactory)						
n	6	27	52	45	2	132
Tot. %	0.71	3.19	6.15	5.32	0.24	15.60
Col.%	100.00	90.00	47.27	9.72	0.84	
2 (Partially proficient)						
n	0	2	52	205	37	296
Tot. %	0.00	0.24	6.15	24.23	4.37	34.99
Col.%	0.00	6.67	47.27	44.28	15.61	
3 (Proficient)						
n	0	1	6	213	189	409
Tot. %	0.00	0.12	0.71	25.18	22.34	48.35
Col.%	0.00	3.33	5.45	46.00	79.75	
4 (Advanced)						
n	0	0	0	0	9	9
Tot. %	0.00	0.00	0.00	0.00	1.06	1.06
Col.%	0.00	0.00	0.00	0.00	3.80	
Total						
n	6	30	110	463	237	846
Tot. %	0.71	3.55	13.00	54.73	28.01	100.00

Table 17 presents the cross frequency table of ELP levels (columns) against reading proficiency categories (rows), which are the two major criteria for ELL reclassification from the state guidelines. Despite the state guidelines, which state that exiting ELL students should be at the highest ELP level (Level 5), students who had an ELP level of 5 upon exiting comprise only 28% of the students reclassified at Grade 6. Most of these students who had an ELP level of 5 upon exiting (80%) scored a “proficient” level in the state reading assessment, which is one level above the suggested level for the content reading assessment according to the state guidelines.

In fact, a majority of students exit one level below the suggested ELP level in state guidelines (Level 4), which is 55% of the students reclassified at Grade 5. Among these students (i.e., even when they exit one level below the suggested ELP level), about half scored at a “proficient” level in the state’s content reading assessment, while the other half scored at a “partially proficient” level, which is the suggested level in content reading assessment according to the state guidelines. This clearly demonstrates that, for many ELL students, meeting the reclassification criteria for the ELP assessment (Level 5) tend to be far more difficult than meeting the criteria for the content area reading assessment (“partially proficient”).

Even with an ELP level of 3, a substantial proportion of students exit. For these students, about half were “proficient” or “partially proficient” in content reading, which is at or above the level suggested by state guidelines. However, the other half of the students scored in the “unsatisfactory” level, which is the lowest level in the content reading assessment. There are a small number of students who exit with the ELP levels of 1 or 2. Not surprisingly, these students mostly had the lowest proficiency level (i.e., the “unsatisfactory” category) in the content assessment. These students might have been reclassified due to some unusual conditions (e.g., some small schools may not have ELL programs), or the reclassification of these students may be due to caveats from our program to identify students who are reclassified. As acknowledged earlier, depending on how the definition of reclassified ELLs is operationalized in our programs, the identification of reclassified ELLs may vary to a slight extent.

As shown in the 20 cells (5 ELP levels x 4 content reading proficiency levels) of Table 17, the most frequent group among students reclassified at Grade 5 are students at ELP Level 4 and at the “proficient” reading proficiency level (25%); the second most frequent group among students reclassified at Grade 5 are students at ELP Level 4 and the “partially proficient” reading proficiency category (24%); and the third most frequent group among students reclassified at Grade 5 is an ELP Level 5 and the “proficient” reading proficiency category (22%). Thus, these three cells comprise a majority (72%) of the reclassified students at Grade 6.

In terms of examining the relationships between ELP levels and post-reclassification growth, we are particularly interested in whether the highest ELP level suggested in the state guidelines relates to students' academic growth subsequent to reclassification. The model we ran for this can be readily extended from Model 2, which we refer to as Model 4. Model 4 basically includes the same variables as Model 2, which examines correlates of growth and additionally includes a binary indicator of whether students exited with the highest ELP level (i.e., Level 5). Model 4 drops three variables that were included in Model 2, which are two indicators of reclassification at different grades and the Native American group. This is because we use a reduced sample of students reclassified only at Grade 6 (N=905) which has no Native Americans in this analysis.

Table 18 presents the results from Model 4, which includes only the fixed effects results for brevity of presentation. Variances and covariances of the growth parameters as well as temporal residual variances did not show any notably different patterns compared to the results from Model 2.

Table 18

Excerpts from Results from Model 4

Fixed effects	Reading			Math		
	Coefficient	*SE*	*p*-value	Coefficient	*SE*	*p*-value
Model for status at Grade 5						
Intercept	574.85	(9.95)	0.00	482.48	(10.72)	0.00
ASIAN	10.32	(11.41)	0.37	24.43	(12.29)	0.05
BLACK	3.27	(24.12)	0.89	11.15	(26.01)	0.67
HISPANIC	-12.78	(9.87)	0.20	-14.86	(10.63)	0.16
LOWSES	-5.13	(4.87)	0.29	-1.72	(5.25)	0.74
CELA5	56.26	(4.26)	0.00	58.38	(4.55)	0.00
Model for growth rate during elementary						
Intercept	27.63	(4.67)	0.00	-	-	-
ASIAN	5.74	(5.34)	0.28	-	-	-
BLACK	2.08	(11.03)	0.85	-	-	-
HISPANIC	1.44	(4.63)	0.76	-	-	-
LOWSES	-1.14	(2.24)	0.61	-	-	-
CELA5	6.55	(1.96)	0.00	-	-	-
Model for growth rate during middle, post-reclassification						
Intercept	17.46	(2.42)	0.00	21.62	(2.32)	0.00
ASIAN	0.32	(2.23)	0.89	1.26	(2.44)	0.61
BLACK	-0.31	(4.64)	0.95	3.19	(5.07)	0.53
HISPANIC	-1.49	(1.92)	0.44	-0.97	(2.10)	0.65
LOWSES	-0.67	(0.96)	0.49	-0.27	(1.05)	0.80
CELA5	-0.66	(0.98)	0.50	-2.60	(1.08)	0.02
Status at Grade 5	-1.11	(0.09)	0.00	-0.53	(0.09)	0.00

Key parameters are the coefficients of the indicator variable when exiting with an ELP level of 5. Students exiting with the highest ELP level, as suggested in the state guidelines, tend to grow significantly more rapidly before reclassification (in reading). While the average annual growth is 28 points, these students tend to grow more rapidly by 6.6 points annually. They also tend to exit at appreciably higher levels by 56 points in reading and by 58 points in math.

The results diverge between reading and math when we turn to the growth rates after reclassification, which is the primary interest in this study. Students exiting with the highest ELP level tend to grow at a similar rate in reading as compared to students exiting with lower ELP levels, holding constant performance status before reclassification (estimate = -0.7, *p*-value = 0.5). However, in math, students exiting with the highest ELP level, as suggested in the state guidelines, tend to grow at a significantly slower rate as compared to students exiting with lower ELP levels, holding constant performance status before reclassification (estimate = -2.6, *p*-value = 0.02).

Conclusion and Discussion

Literature on ELLs has discussed the vast prevalence of long-term ELLs within the ELL population (Grissom, 2004; Mitchell, Destino, & Karam, 1997; Parrish et al., 2006). While such literature well depicts the potentially detrimental status of long-term ELL students in public education, it is important to note that a good proportion of ELL students do exit from ELL status or get reclassified as fully proficient in English and are mainstreamed by the time they finish elementary school. This present study focused on such ELL students who were reclassified around the time they finished elementary school (specifically students reclassified at Grades 4, 5, or 6) and attempted to examine whether the reclassification decisions used for these students are valid and supportive of their subsequent learning. With valid reclassification decisions, the reclassified students will continue to grow in their academic performance throughout the grades in mainstream classrooms. Conversely, students who exited with improper reclassification decisions may struggle and may not grow adequately. In fact, these students may eventually reemerge as ELLs in later grades.

One set of analyses in this paper estimates growth rates after reclassification and compares them to growth rates of the other students over the same period and thereby attempts to draw inferences about the existing reclassification decisions. The general trend both in reading and math indicates more rapid average growth rates of reclassified ELLs than non-ELLs holding constant prior academic status, which means that the reclassified ELLs tend to catch up with their non-ELL peers over the grades. This pattern may suggest that the existing reclassification decisions were on average supportive of ELL student learning. However, when their average trajectories are compared to state-designated academic proficiency levels, there is little evidence that ELLs are catching up relative to their proficiency classifications over the grades: the initial gaps, either minor or sizeable, tend to persist over time. Thus, from these findings and trends we can draw the following conclusions with more certainty:

First, although there is evidence that reclassified ELLs tend to continue to catch up to their non-ELL peers after reclassification, the magnitudes may be very modest in virtual scales over the grades and insufficient to attain proficiency. Secondly, there is no evidence of ELLs falling behind in academic growth after reclassification, either relative to their non-ELL peers or in terms of absolute academic proficiency levels, which may suggest that reclassification decisions on average did not hamper ELLs' subsequent academic growth in a state where reclassification decisions are made locally (i.e., delegated to districts or schools). These findings provide positive empirical evidence to the validity of the existing reclassification system in the framework of this study.

As noted earlier, the analyses reported here are limited from a causal perspective. Thus, we carefully make comparisons and interpret the findings above but do not infer that more rapid growth is the "effect" of reclassification. This is because the comparisons are based on non-equivalent groups with potentially very large magnitudes of differences in many preexisting characteristics. Even though the analyses reported in this paper controlled for prior performance status, they do not likely explain away all the preexisting differences embedded in these intact groups. In addition, as noted earlier, we opted not to depend on the ITS design in this present study. Within a growth modeling framework, it was not feasible to compare growth rates of the same reclassified students before and after reclassification based on the ITS design because natural growth was discontinuous regardless of reclassification and because the time series from annual assessments was not long enough. Although there may be available methods to obtain estimates of causal effect under certain sets of assumptions, such as fixed effect models (see, e.g., Bifulco & Ladd, 2006; Rivkin, Hanushek, & Kain, 2005), such models may be limited in what growth modeling techniques can do, such as estimating individual growth over time, investigating correlates of change, and incorporating data nesting structure when necessary (see Raudenbush, 2009 for more details). This present study chose to utilize growth modeling techniques that incorporate random effects (see Diggle, Liang, & Zeger, 1994; Raudenbush & Bryk, 2002, Chapter 6; Singer & Willet, 2003, Chapter 3) and focused on investigating growth trajectories of various intact groups and their association with key variables of interest that can address our research questions. Therefore, the results are inconclusive at this time with respect to overall positive or negative causal effects of reclassification.

Use of growth modeling techniques enabled the estimation of inter-individual variability as well as average trajectories, which indicated great heterogeneity in how students grow, as well as where students are in academic performance. For example, a reclassified ELL student who starts out at a similar level to other reclassified ELL students, but had a growth rate of 1 *SD* above the average growth rate, could either catch up to or outperform non-ELL peers by the end of Grade

8. Also, a reclassified ELL student with a growth rate of 2 *SD*s below the average growth rate could perform even lower than where they started at in Grade 5, which means that their learning is negative, or so minimal that their academic proficiency level is assessed at a level that does not even retain the level of knowledge from previous grades.

Such a great extent of individual differences among reclassified ELLs in how they grow over the middle school years naturally leads to a question of correlates of growth/change: what factors would explain subsequent success of reclassified ELL students? The present study examined student and district factors but found that state data on student demographics was of little value in predicting change after reclassification. Among the various student characteristics, the only significant correlate of change was the ethnicity category. Consistently in both reading and math, Hispanic reclassified ELL students tended to grow significantly slower than the other reclassified ELL students. As the ELL population consists of 87 % of Hispanic students, we would like to learn more about underlying factors that may explain the difference between Hispanic students and students of other ethnicities and explain the within-group heterogeneity of Hispanic ELL students.

In addition, several districts that showed significantly more rapid change rates had a very small number of reclassified ELLs, with an exception of one district. We anticipated seeing more variability across districts given that the data is from a local control state in which districts use their own criteria under state guidelines and make reclassification decisions, but that did not turn out to be the case.

Grades, at which reclassified ELL students exit, are predictive of performance level. Increasing and statistically, gaps with non-ELL students exist in performance status/levels upon exiting as we go from students reclassified at Grades 4 to 5 and also from students reclassified at Grades 5 to 6. But again, reclassified grades did not emerge as a significant predictor of growth subsequent to reclassification. This finding may suggest that whether ELL students are reclassified earlier or later might not matter much, on average, in terms of how they grow in or benefit from mainstream classrooms over time. Caution is needed to generalize this finding, because we are examining only three adjacent grades. Also, we noted varying practices around reclassification. For example, depending on the schools or districts, students reclassified in the 4^{th} grade may have received similar instruction and stayed in similar class settings to students reclassified in the 5^{th} grade.

Lastly, this paper examined components of existing reclassification criteria, by examining the relationship of ELP levels upon exiting to post-reclassification growth. In our framework, use of major components of the reclassification criteria is valid if the efficiency of the components determines students' readiness for mainstream classrooms. Contrary to state guidelines on ELL reclassification, a majority of ELL students were able to exit ELL status at an ELP level of 4 or below instead of the highest ELP level of 5. After level of content area achievement at time of exiting was controlled for, students who were reclassified with the highest ELP level (i.e., Level 5) did not show any significant difference from other ELL students who were reclassified at a lower ELP level in terms of subsequent learning rates in reading, whereas they showed significantly slower learning rates in math. This may suggest that too stringent ELP criteria may not be useful in ELLs' subsequent learning in mainstream classrooms. Furthermore, this finding may suggest that, in subjects like math in which a sequence of learning is especially important and language is less required, prolonged ELL status due to too stringent ELP criteria may be detrimental to learning subsequent to reclassification in mainstream classrooms.

We do not have analyses or results to explain the relationships between ELP level and subsequent performance. However, it is notable that this set of analyses used the sample of ELL students reclassified at Grade 6. Grade 6 can be considered as a time when students begin to be assigned to classes based on ability tracking. Additionally, the Grade 6 curriculum starts to build up math knowledge for core math classes that are critical to high school graduation and entrance to post-secondary education in later years (Hakansson & Woods, 2009). Thus, in cases where Grade 5 ELL students who are academically ready for mainstream classrooms are retained as ELLs in the Grade 6 due to too stringent ELP criteria, they may miss the opportunity to take more competitive math classes that will build prior knowledge for subsequent years. Missing the opportunity to build up prior knowledge on time may keep them from learning as rapidly as students who are reclassified earlier and receive the opportunity to be in a class that corresponds to their math ability on time. In such cases, waiting for higher ELP levels to reclassify students may come at the cost of missing out on the opportunity to build on core academic knowledge.

A large portion of individual differences in growth among reclassified ELL students remains unexplained, which suggests the need to collect data on additional variables that are more relevant to ELL population at both the student and local levels. Individual differences in academic growth may be partly explained by other student characteristics that are more relevant to ELL population, such as the age of entry to the United States, previous schooling experience in the States, prior schooling experience in their country, and literacy levels in their native language, to list but a few. For example, the heterogeneity of ELL students at the high school level is often noted, including long-term ELLs, recently-arrived and highly-educated students, and recently-arrived and under-educated students (Freeman & Freeman, 2007; Olsen, & Jaramillo, 1999). Students from each of these groups may be expected to be distinctively different in their academic growth over grades as well as being distinct in many other characteristics, but variables available from the state data systems are usually too rough to contain such detailed information on student background. For instance, eligibility for free or reduced lunch usually serves as a proxy for socioeconomic status (SES) in the entire sample. However, this indicator may not well serve the subgroup of ELL students, since a vast majority of ELL students are receiving free or reduced lunch. To understand the SES of the ELL students, one may need a more fine-grained measure.

Differences in academic growth of ELL students may also be explained by school-level characteristics, practices, or policies. Academic growth of ELL students could be related to contextual factors such as concentration of ELL students and/or of low SES students, school practices such as the quality and availability of ELL programs and teachers in schools, or reclassification criteria schools use. There is some amount of evidence as to the association of school practices and ELLs' learning. In a large statewide study contrasting practices in demographically similar schools that were relatively more or less effective in promoting ELL learning, EdSource (2007) found four broad practices associated with effective schools: using assessment data to improve instruction and achievement; ensuring availability and adequacy of instructional resources; prioritizing learning objectives and monitoring progress; and implementing coherent, standards-based curriculum. Among the specific practices differentiating effective schools for ELLs was the use of recent ELD programs. Similarly, Parrish and colleagues (2006) used schools with relatively high and low reclassification rates to identify factors critical to redesignation. Identified factors included staff capacity to address ELL needs, schoolwide focus on ELD and standards-based instruction, shared priorities and expectations within and across grades, data-based decision making, and systematic, ongoing assessment. Schools that showed relatively high rates of redesignation, moreover, use carefully designed plans for ELD services to ensure that academic language and literacy development was fostered across the curriculum and that there was sustained professional development and technical assistance to support ELD practices.

The last point in school-level factors – reclassification criteria schools use – is closely related to our study. One set of our analysis assumed variability across districts. However, in reality, the actual agents that make the reclassification decisions may have been individual schools instead of districts. In such cases, the relationships of reclassification to subsequent growth rates may vary across schools rather than districts.

The state examined in this present study requires ELL students to take annual assessments of academic proficiency and ELP and use both assessments as major sources for reclassification decisions. This study presents an interesting finding that could shed light on ELL reclassification criteria: too stringent ELP criteria for reclassification may hinder students’ subsequent learning in mathematics. However, this finding may raise more questions rather than resolving issues about optimal combination of reclassification criteria. For example, most states, including the one included in the present study, use a conjunctive rule for reclassification. ELLs must meet minimum criteria on several indicators and failing one stops ELL students from exiting. But what if states take a more differentiated approach in combining on academic proficiency and on ELP? How might they weigh information on academic proficiency and information on English proficiency? Should the weights for the two types of assessments and other sources of information be the same across different settings or across school levels?

Some of the questions raised above, with regard to correlates of subsequent growth and more optimal reclassification criteria, will be answered by the next phase of this study. A subsequent goal of the study will examine more directly the reclassification criteria and the ELL program and service features that are associated with districts or schools that foster higher versus lower growth. In addition to the extant state data, we will collect information about reclassification criteria from the 38 school districts in State A that serve the highest concentration of ELLs (the study sample of this report). Through telephone interviews with these districts, we will query what sources of information the district uses (e.g., ELD scores, state content assessments, teacher judgments, local measure, others), the criterion level needed for each source (e.g., ELP level, proficiency or other performance levels on state content tests), and how information within and across sources is combined. Information will be quantified and coded in summary variables (such as stringency of criteria, stringency of required ELD performance, and specific ways of combining information across different sources) and used in quantitative analyses in order to investigate how differences in reclassification criteria relate to differences in relative success in promoting subsequent student achievement.

References

Bifulco, R., & Ladd, H. F. (2006). The impacts of charter schools on student achievement: Evidence from North Carolina. *Education Finance and Policy, 1*(1), 50-99.

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control.* San Francisco, CA: Holden-Day.

Callahan, R. (2005). Tracking and high school English learners: Limiting opportunity to learn. *American Educational Research Journal, 42*(2), 305–328.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Choi, K., & Seltzer, M. (2010). Modeling heterogeneity in relationships between initial status and rates of change: Treating latent variable regression coefficients as random coefficients in a three-level hierarchical model. *Journal of Educational and Behavioral Statistics 35*(1), 54–91.

Cummins, J. (1980). The entry and exit fallacy in bilingual education. *NABE Journal 4*(3), 25–59.

Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics, 2*(2), 132–149.

Diggle P. J., Liang K.Y., & Zeger S. L. (1994). *The analysis of longitudinal data*. New York: Oxford University Press.

EdSource (2007, September). *Summary report: Similar English learner students, different results*. Palo Alto, CA: Author.

Escamilla, K., Mahon, E., Riley-Bernal, H., & Rutledge, D. (2001). *Limited English proficient students and the Colorado student assessment program (CSAP): The state of the state, year II.* Denver, CO: Colorado Association for Bilingual Education.

Freeman, Y., & Freeman, D. (2007). Four keys for school success for elementary English learners. In J. Cummins,. & C. Davison. (Eds.), *International handbook of English language teaching*. New York: Springer.

Gándara, P., Gutierrez, D., & O'Hara, S. (2001). Planning for the future in rural and urban high schools. *Journal of Education for Students Placed At Risk, 6*(1–2).

Grissom, J. B. (2004). Reclassification of English language learners. *Education Policy Analysis Archives, 12*(36), Retrieved August 2008 from http://epaa.asu.edu/epaa/v12n36

Hakansson, S. W., & Woods, K. (2009, November 6). Supporting mathematical proficiency for all students. *California Mathematics Council - South 50th Annual Math Conference* Retrieved June 13, 2010, from www.cmpso.org

Harklau, L. (2002). ESL versus mainstream classes: Contrasting L2 learning environments. In V. Zamel & R. Spack (Eds.). *Enriching ESOL pedagogy: Readings and activities for engagement, reflection, and inquiry.* Mahwah, NJ: Lawrence Erlbaum Associates.

Kim, J., & Herman, J. L. (2008). Investigating ELL assessment and accommodation practices using state data. *Providing validity evidence to improve the assessment of English language learners* (CRESST Report 738). Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing (CRESST).

Kim, J., & Herman, J., L. (2009). **Error! Hyperlink reference not valid.**. ***Error! Hyperlink reference not valid.***, ***Error! Hyperlink reference not valid.*** (**Error! Hyperlink reference not valid.**), 212 – 231.

Linquanti, R. (2001). *The redesignation dilemma: Challenges and choices in fostering meaningful accountability for English learners* (Policy Report No. 2001-1). Santa Barbara, CA: University of California Linguistic Minority Research Institute.

Maxwell-Jolly, J., Gándara, P., & Méndez Benavídez, L. (2007). *Promoting academic literacy among secondary English language learners: A synthesis of research and practice* (Policy Report). Oakland, CA: University of California, Davis, School of Education, Linguistic Minority Research Institute.

Mitchell, D. E., Destino, T., & Karam, R. (1997). *Evaluation of English language development programs in the Santa Ana Unified School District: A report on data system reliability and statistical modeling of program impacts.* University of California, Riverside School of Education: California Educational Research Cooperative.

Muthen, B., & Curran, P. (1997). General growth modeling with interventions: a latent variable framework for analysis and power estimation. *Psychological Methods, 2,* 371-402.

Olsen, L., & Jaramillo, A. (1999). Turning the tides of exclusion: A guide for educators and advocates for immigrant students. Oakland, CA: California Tomorrow.

Parrish, T., Perez, M., Merickel, A., & Linquanti, R. (2006). *Effects of the implementation of Proposition 227 on the education of English learners, K-12: Findings from a five year evaluation* (Final Report). Palo Alto, CA and San Francisco, CA: American Institutes for Research and WestEd.

Raudenbush, S. W. (2009). Adaptive centering with random effects: An alternative to the fixed effects model for studying time−varying treatments in school settings. *Journal of Education, Finance and Policy, 4*(4), 468 – 491.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods,* (2nd ed.). Newbury Park, CA: Sage.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement, *Econometrica, 73*(2), 417-458. SAS Institute Inc., (2002–2005). SAS Online Doc 9.1.3, [Computer Software]. Cary, NC: Author.

Seltzer, M., Choi, K., & Thum, Y. M. (2003). Examining relationships between where students start and how rapidly they progress: Using new developments in growth modeling to gain insight into the distribution of achievement within schools. *Educational Evaluation and Policy Analysis, 25*(3), 263–286.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.

Silver, D., Saunders, M., & Zarate, E. (2008). *What factors predict high school graduation in the Los Angeles Unified School District?* (California Dropout Research Project Report No. 14). Santa Barbara, CA: University of California, Linguistic Minority Research Institute.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Methods for studying change and event occurrence.* New York, NY: Oxford University Press.

State A's Department of Education (2007). *State A's Student Assessment Program: Technical Report*.

Watt, D., & Roessingh, H. (1994). ESL dropout: The myth of educational equity. *Alberta Journal of Educational Research, 40,* 283–296.

Wolf, K. M., Kao, J. C., Herman, J. L., Bachman, L. F., Bailey, A., Bachman, P. L., et al. (2008). *Issues in Assessing English Language Learners: English Language Proficiency Measures and Accommodation Uses–Literature Review* (CRESST Report 731). Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing (CRESST).