CRESST REPORT 818

UNDERSTANDING PATTERNS AND PRECURSORS OF ELL SUCCESS SUBSEQUENT TO RECLASSIFICATION

AUGUST, 2012

Jinok Kim

Joan L. Herman

National Center for Research
on Evaluation, Standards, & Student Testing

UCLA | Graduate School of Education & Information Studies

Understanding Patterns and Precursors of ELL Success Subsequent to Reclassification

CRESST Report 818

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

August 2012

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 R305A09058101, 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. (2012). *Understanding patterns and precursors of ELL success subsequent to reclassification* (CRESST Report 818). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

TABLE OF CONTENTS

UNDERSTANDING PATTERNS AND PRECURSORS OF ELL SUCCESS SUBSEQUENT TO RECLASSIFICATION

Jinok Kim and 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 present study focused on 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. In doing so, this paper also explores methods that allow for drawing sound inferences on student learning subsequent to reclassification. Recent advances in growth modeling are drawn upon to make comparisons in subsequent learning more meaningful. The study found that 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 scale values over the grades and insufficient to attain proficiency. The study also found that there was no evidence of former ELLs falling behind in academic growth after reclassification, either relative to their non-ELL peers or in terms of absolute academic proficiency levels.

Introduction

Reclassification is a key milestone for English language learners (ELLs). Reclassification is the point when ELL 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. However, the validity of existing criteria and procedures lack an empirical base; in fact, reclassification practices are formulated and implemented with little knowledge of the factors that may influence their success.

The tension between assuring that students have sufficient English language proficiency (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. 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 (see Parrish, Perez, Merickel, & Linquanti, 2006). The cumulative effects of diminished access to academic coursework over time can be significant, potentially preventing ELLs from entering postsecondary 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).

In beginning to resolve this dilemma, our study takes a particular view on the meaning of effective reclassification: it can and should be judged by its consequences. As one of the most immediate consequences, students' ability to benefit should be evidenced in subsequent outcomes—such as academic performance on state tests of reading and mathematics. With valid reclassification decisions, the reclassified students will continue to grow in their academic performance in mainstream classrooms. Conversely, students who exited with improper reclassification decisions may not grow adequately and may eventually reemerge as ELLs in later grades. Based on such a perspective, we assess the validity of existing systems (a) in terms of gross consequences of reclassification—or the subsequent academic success or failure of reclassified ELLs in mainstream classrooms; and (b) by examining differences in reclassification criteria as well as in various student and district factors related to differences in relative success in promoting subsequent student achievement. We focus on ELL students who were reclassified around the time they finished elementary school (specifically students reclassified at Grades 4, 5, or 6) and examine whether the reclassification decisions used for these students are valid and supportive of their subsequent learning. In doing so, this paper also explores methods that allow for drawing sound inferences on student learning subsequent to reclassification. As will be seen, we draw on recent advances in growth modeling to make comparisons in subsequent learning more meaningful.

Background and Context

Issues and Research Concerning ELL Reclassification

In addressing issues concerning ELLs' reclassification, we first review states' and local agencies' current policies and practices regarding reclassification decisions. We then consider more basic research about the expected time it takes for non-native speakers to acquire sufficient English proficiency for schooling and contrast these estimates with current realities. Lastly, we review prior research that contributes to our work and lay out research questions of this study.

Current status of reclassification criteria. The optimal time to place ELLs in mainstream English classrooms remains a highly controversial issue, as exemplified by California's Proposition 227, an initiative requiring all ELL students to be mainstreamed and taught overwhelmingly in English after a maximum one year transition period. Opponents and proponents argued vociferously to advance their views; then, each side used available data to claim success—when the reality was far from clear (see Unz, 1997; and for critics, Gándara, 2000 and Mora, 2000). Because the research basis for making mainstreaming or reclassification decisions remains slim, it may not be surprising that criteria for reclassifying students from ELL to Reclassified as Fluent English Proficient (RFEP) status vary substantially across states, as documented by a recent report reviewing statewide practices related to ELLs. Of the 48 states in which the information was obtained, 12 used only the results of their ELD test for purposes of reclassification; 7 used both ELP and state content-area tests in some combination; and in 17 states, districts were in charge of reclassification of ELLs. While these findings suggest that ELP, as measured by state-chosen ELD assessments, are a primary criterion for reclassification in almost all states, even the use of this common criterion can mask substantial variation. States use different ELP tests, which are not comparable, and even for states using the same test, ELP level that students must meet for reclassification can vary. Within-state variation also appears considerable. As noted earlier, in the 17 states with no statewide criteria, the reclassification decisions are left to the discretion of schools or local education agencies, adding substantial within-state variation to the decision making process. For example, in California, which has the largest population of ELLs, the substantial variability in reclassification rates across districts has been repeatedly reported by a variety of sources (see Abedi, 2008; Jepsen & de Alth, 2005; Linquanti, 2001; Parrish et al., 2006). For example, while all use the results of the state's ELD measure, districts vary in the overall all level of proficiency required for redesignation (for example, requiring a level 4 versus a level 5) as well as how they treat the component ELD scores (e.g., reading, writing, speaking, and listening). Similarly, all districts use results of the state wide reading test, but for some, the criterion may be set at the 35th percentile; while for others it is set at the 50th, and a variety of other sources of information may be included in the decision, including teacher judgments, the results of idiosyncratic local measures, and parent input.

How long does it take for non-native speakers to acquire English language proficiency (ELP) for schooling? One fundamental and critical issue that should underlie policies and practices about reclassification of ELLs is the time needed for non-native speakers to acquire second language proficiency sufficient for schooling. With available theory and research, ELD instructional planning and reclassification can be built on reasonable expectations. Cummins

(1981) found that it takes immigrants two to three years to acquire basic communication skills in foreign language – e.g., required to navigate social situations – but was adamant in pointing out that skill in basic social communication was insufficient for school instruction. Based on re-analyses of large data bases from Canada that include 1200 immigrants in Grades 5, 7, and 9, and an examination of the relationship between the time it takes to reach the 50th percentile in various tests on English skills, and length of residence and age of arrival, Cummins found that a period of 5 to 7 years is needed to reach proficiency in order to reach native-speaker levels in school language. With slight variations, these findings are generally confirmed by other research in various contexts (see, e.g., Collier, 1987, 1989; Klemser, 1993; Hakuta, Butler, & Witt, 2000). But the collective findings also show that context matters. For example, Collier (1987) found that it should take students below age 12, adequately schooled both in their primary language and in second language, from 5-7 years to reach national norms on standardized tests in reading, language arts, social studies, and science, and as little as 2 years in math. In contrast, young students who had immigrated at ages 4-6 years, and thus had little or no formal schooling in their primary language, tended not to reach the 50th percentile in 6 years, and it was projected to take much longer (7-10 years).

How long does it take for ELLs to get Redesignated Fluent English Proficient (RFEP) status? With some obvious common ground, how long it actually takes for ELLs to get RFEP status is a fairly distinct question from how long second language acquisition usually takes. First, the answer depends on reclassification criteria as currently adopted and implemented by state or local educational agencies, which tend to be inconsistent and potentially ambiguous within and across states, as noted earlier. Second, research on language acquisition suggests that the answer may vary depending on the nature and heterogeneity of the ELL population studied, (e.g., the ELL population currently in the U.S. public education system may be different than the samples that have been used for studies on second language acquisition.). Mitchell, Destino, and Karam (1997) use Santa Ana district ELL data in California with survival analysis (also known as event history analysis) statistical techniques to estimate time to redesignation. They conclude that it takes approximately 10.6 years for an ELL who starts from the lowest level to reach the highest level (i.e., "Redesignated as FEP"); in so doing they also clearly note that using other, more naïve techniques may seriously underestimate expected time duration by ignoring students who have not achieved proficiency by the end of the study (such cases are termed "censored" observations in survival analysis literature). In a more recent large-scale study using survival analysis techniques with data from the entire California state, Parrish, et al. (2006) find that there is less than a 40% probability of ELLs being redesignated in 10 years, and an estimated 75% of

ELLs remain in that status after 5 years in California, which reconfirmed the findings by Grissom (2004).

At what point should ELLs get RFEP status in ELL policies and practices? As noted in the introduction, when ELLs should be reclassified to fully engage with mainstream classrooms has been a controversial issue in which opposite arguments have been supported by research literature. On one hand, premature exit is strongly opposed. As Cummins noted, an "exit fallacy" is deeply ingrained in policy, reflecting the assumption that "mainstreaming minority children out of a bilingual program into an English-only program will promote the development of English literacy skills more effectively than if children were maintained in a bilingual program" (Cummins, 1980, p. 49). Theorists advocating against the "exit fallacy" argue that it takes students much more time to gain the proficiency needed for schooling – academic language – than to gain basic oral proficiency that may provide the illusion of proficiency. While newer measures of English proficiency that address the development of academic language proficiency mitigate the problem of students being reclassified based on social rather than academic language (for example, see Wolf, Kao, Herman, et al., 2008; Abedi, 2003), the basic problem of premature exit may remain: ELLs with insufficient ELP may mistakenly be reclassified, mainstreamed, and get no instructional support for their English, and as a result turn out as low-performing students.

On the other hand, although not fully on point to the issue of when to exit/reclassify ELLs, another body of research underscores potential negative consequences of prolonged ELL status, especially long-term ELL designation in secondary schools. The theory behind this criticism is that poor performance may not only be due to limited English proficiency, but also due to poor academic preparation, or a combination of the two (Callahan, Wilkinson, Muller, & Frisco, 2008; Lam, 1993). The authors argue that current ELL policies and practices, for example, "preference given to English acquisition over academic training, coupled with organizational constraints inherent in ensuring the delivery of linguistic services required by law," may preclude students' access to challenging academic coursework, which in turn may keep them from having the academic preparation necessary for entry into higher education (Callahan et al., 2008, p.3). Practitioners also reveal similar conflicting viewpoints regarding early versus later exits. Based on interviews with district administrators represent both extremes of high and low redesignation rates. Parrish and colleagues (2006) summarize the tensions inherent in practice: "[English Learners] redesignated prematurely may lose needed instructional services and be placed at greater risk of educational failure, while long-term ELs often face segregated track placement and reduced access to courses needed for post-secondary education. (p. V-23)." The inherent

dilemma in current ELL redesignation policy and practice provides a prime rationale for the proposed study.

Potential underlying sources of gaps in time to reclassification. Analysis of previous studies relevant to reclassification of ELLs, including the ones referred to or cited earlier, reveal the complexity both of reclassification policy, and practice of disentangling the factors that may influence its success. Previous studies cited earlier have identified potential problems in current reclassification, qualitatively analyzed criteria, and student characteristics that may relate to high versus low redesignation rates, and examined related research questions, such as how long it takes for non native speakers to acquire ELP or be reclassified; but none of the existing literature has directly dealt with reclassification systems and their consequences, and more specifically with the consequences of various reclassification criteria. Moreover, while some studies have examined relationships between ELD strategies and program types; and subsequent performance on state tests relative to reclassification outcomes (see, for example, Edsource, 2007; Parrish et al., 2006; Ramirez, Yeun, & Ramey, 1991; Thomas & Collier, 2002; Rossell & Baker, 1996, Slavin & Cheung, 2003, for meta analysis; and de Cos, 1999, for a critical review), none have linked ELD success to both reclassification and subsequent performance. These are the missing links which the current study seeks to address in a rigorous empirical study that directly addresses reclassification, strongly in terms of internal validity (when relevant) and external validity. For example, as noted earlier, research is fairly consistent on how long it takes for ELLs to achieve ELP. Empirical evidence from studies of thousands of immigrants, both in Canada and U.S. contexts, show that it generally takes 4-7 years for ELLs to acquire ELP as needed for schooling. Yet these research findings are in sharp contrast to data from current educational practices. As noted earlier, in California, after 5 years of ELL designation, less than 25% get reclassified to RFEP, and after 10 years, less than 40% get reclassified to RFEP.

Then, what might explain this significant gap between research and practice, e.g., the apparent 60% of ELLs who cannot be reclassified in 10 years of schooling? One quick explanation may be important differences in student demographics between the studies by Cummins, Collier, and others in the 1980s or before, and those who are in the current U.S. public school system. Studies by Collier (1987) cited earlier, for example, intentionally excluded older students who had no formal education in their primary language. Further, the population in the current U.S. public system may be more heterogeneous—including more students in extreme poverty, students with disabilities, students who have minimal proficiency even in their primary language and/or may be substantially different in their entering ELD proficiency, native language and ethnicity; all of these characteristics have shown significant relationship to ELLs' performance and to reclassification (Parrish, et al., 2006; Abedi, Leon, & Mirocha, 2003; Abedi,

2008; Kim & Herman, 2008). Little empirical research directly addresses how much of the gap in time is attributable to such demographic differences, which warrants longitudinal studies. Also, it is likely there are other factors than student heterogeneity that underlie the gap in time to reclassification to RFEP for a substantial proportion of ELLs.

Inadequate opportunities to learn for ELLs, including quality of ELD programs and services, may be an additional source in perpetuating ELL status. At the elementary school level, some districts indicate no redesignation until ELLs reach Grade 3 (Parrish et al., 2006, p. V-11, Exhibit V-3), meaning that once initially designated on the basis of home language and ELP performance, all ELLs remain in that status for at least four years (Kindergarten through Grade 3), and the quality of ELD and instructional services for ELLs may be uneven. Research, for example, shows that ELL students are more likely than non-ELL students to have inexperienced and unqualified teachers (Gándara & Méndez Benavídez, 2007) and significant variation in programmatic elements associated with ELL success. For example, in a large state wide study contrasting practices in demographically similar schools that were relatively more and 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 et al. (2006) used schools with relatively high and low reclassification rates to identify factors critical to redesignation. Identified factors included staff capacity to address EL needs, school wide focus on ELD and standards-based instruction; shared priorities and expectations within and across grades; and systematic, ongoing assessment and data-based decision making. 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. These studies suggest programmatic features that can be used in the current study to help explain ELL success subsequent to reclassification.

Lastly, one very plausible source underlying the gap in time to reclassification to RFEP may be the reclassification criteria themselves as currently adopted and implemented by state and local educational agencies. For example, some criteria may be overly protective against prematurely reclassifying ELLs to RFEP, which result in holding back ELLs who are ready for challenges in mainstream classrooms. An empirical study of statewide reclassification criteria (Kim & Herman, 2008) found that in one state which uses a uniform and single criterion for

reclassification (i.e., proficiency in ELP as measured by the state ELD assessment in four modalities), ELLs in Grade 7 who met proficiency in the writing assessment on the state wide assessment for all student (thus primarily designed for non-ELL students) tended not to achieve proficiency on the state's ELD writing assessment, suggesting that the standard for reclassification to RFEP was higher than that for statewide standards for all students. Further, the Parrish et al. (2006) study found that districts with very low reclassification rates tended to use grades from multiple local tests in addition to multiple criteria suggested by state, and/or to set higher cut scores for required ELP levels or for required state assessment scores. While evidence of the relationship between stringency of reclassification criteria and subsequent performance is light (and a motivation for the current study), Kim and Herman's (2008) findings in a cross sectional study of three states are suggestive. The performance of recently reclassified students (less than two years since reclassification) tended to be higher relative to non-ELLs in the state with the most stringent criteria, while the performance of students who were reclassified more than two years previously, on average, performed higher relative to non-ELLs in all states, regardless of the stringency of reclassification criteria. Given the cross sectional nature of the study, it was not possible to ascertain whether reclassified students in a state with the lenient criteria catch up with their non-ELL peers after more than two years, or whether the findings reflect selection bias due to the characteristics of students who get reclassified in earlier grades (i.e., two or more grades earlier) compared to those in later grades (i.e., recently reclassified students).

The Present Study

Available evidence indicates inconsistencies and ambiguities of reclassification criteria, and few empirical studies have attempted to show whether certain types of reclassification criteria are more desirable than others in success in mainstream classrooms in subsequent years. While success in state annual assessments has been used in studies evaluating the relative effectiveness of various ELD instructional services, subsequent success as measured by state annual assessments has not been linked to the validity of reclassification policies or practices, especially using entire statewide data. Yet success in mainstream classrooms, as measured by annual state assessments and meeting grade-level achievement standards over time, is one of the ultimate goals of reclassification, as well as of ELL education in general, and thus serves as important validity evidence for assessing the effects of reclassification policy.

Further, while many studies have examined the achievement gap between ELLs and non-ELLs using cross-sectional studies (i.e., using state or national assessment outcomes for one year), and continue to confirm ELL students perform dramatically lower than non-ELLs, and that reclassified students tend to close the gap with and even surpass their non-ELL peers (i.e., RFEP

students on average perform better than non-ELL students; GAO, 2006; Kim et al., 2008; Perie et al., 2006), such studies are flawed for the purpose of examining reclassification criteria. The primary reason is the selection bias: the RFEP group intentionally contains the best performing students in the ELL group, those who have met the proficiency and other requirements for reclassification. Especially in comparisons involving only one time point, it is near impossible to connect reclassification criteria and differences in achievement between RFEP students and current ELL students.

The present study aims to fill in such a gap in research and examine the validity of reclassification policies or practices in relation to student achievement. First, using statewide data from multiple years, 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 academic achievement over the years before and after reclassification in the same group of students, this study enables us to lessen the selection bias of the subgroups between reclassified and current ELL students, and draw sound inferences concerning the relationships between ELL reclassification and their achievement.

Second, our research aims to decrease the selection bias further—for example, in comparing intact groups (e.g., comparing reclassified ELL students or other ELL students) - by drawing on the strength of recent advances in growth modeling techniques. These techniques allow 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.

Third, 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 paper goes beyond the average differences to examine for whom, under which criteria, and under which settings, reclassified ELL students receive greater benefits and experience more success. We purposely chose a state with local control to be able to examine a range of reclassification policies and practices that are currently implemented (see the State A's context below), and examine student and district factors as well as individual-level reclassification criteria used, which are available from extant state data.

Lastly, in addition to examining extant data, the present study also incorporates qualitative data about districts' and schools' existing policies and practices regarding reclassification criteria and decision making. With these data, the study explores how reclassification decision-making may be associated with districts that foster higher versus lower growth.

The State A context. A local control state, State A leaves reclassification decision making at the discretion of local school district, but provides districts with suggested guidelines. State suggested 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 also are 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 generally follow state guidelines but may adapt specific criteria used for reclassifying ELL students. The study uses state longitudinal data and date of redesignation to infer which criteria may have been used for individual students. As will be seen, results suggest that districts and schools deviated from state guidelines.

State A also leaves at the discretion of local districts what, if any, services are offered to ELLs for the two years subsequent to reclassification. That is, according to federal guidelines, ELL students continue to count as ELLs and their ELP level is monitored for two years after redesignation. Some schools or districts may continue to offer varying levels of continuing assistance in EL (personal communication with state personnel, 2009). Such varying practices make it difficult to know from existing data the exact timing of transition or full maintstreaming for individual ELLs.

Research Questions

We outline the research questions of the study in the following. 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 demographics of reclassified ELLs who are associated with greater subsequent academic success?
4. How do differences in student performance relative to specific reclassification criteria (e.g., ELP levels) just before reclassification relate to differences in subsequent academic success?
5. To what extent do the estimated growth trajectories of reclassified ELLs vary across districts?

Lastly, we incorporate interview data and examine the following research question:

6. Do districts reclassification criteria predict relative success in promoting subsequent academic success of reclassified ELLs?

Research Design and Methods

This section describes a three phase study incorporating quantitative and qualitative data. Details are provided on data sources and analytic techniques.

Three Phases of the Study

The first phase of the study used State A's extant data (see quantitative data section below) to examine the estimated ELL growth patterns after reclassification and compared them to those of other students. In the second phase of the study, we collected qualitative data through semi-structured telephone interviews to gather information about district and school redesignation practices. In the third phase of the study, using variables from the interview data, we studied the relationships between different reclassification criteria and subsequent academic growth. In doing so, we aim to suggest reclassification criteria that may be premature, optimal, or delayed.

In the second phase of the study, we collected qualitative data through semi-structured telephone interviews (see qualitative data below). The protocol asked about the sources of information the district uses (e.g., ELD scores, state content assessments, teacher judgments, local measure, others) to make redesignation decisions, 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 was quantified and coded in summary variables (such as stringency of required ELD performance and specific ways of combining information across different sources).

The phases were iterative and informed each other. For example, the first phase of the study identified 38 districts as a study sample; and suggested a component in reclassification

criterion that might influence ELLs' subsequent success. This helped us formulate a hypothesis about the types of district reclassification policies that might be more beneficial to ELL students. Based on such a hypothesis, qualitative data were first coded and then combined to create a an overall variable that seemed to differentiate district practices. The district variable was in turn tested in the quantitative analysis using similar methods in the first phase of the study to see whether district ELL reclassification policies and practices was related to relative district success in promoting more rapid growth of reclassified ELLs.

Data Sources

Quantitative data. State A provided six years of longitudinal data on the statewide cohort that started in Grade 3 in 2003–2004, which enabled us to track these students through Grade 8 (in the year 2008–2009). The data include variables such as academic achievement based on state content assessment; 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.

We focus 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) these are the ELL students who are 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. Data across the six years were merged to create ELL status profiles that enabled us to identify individuals reclassified at Grades 4, 5, and 6. This paper omits the procedures we used based on the ELL status profiles, as they are lengthy and described elsewhere (CITE).

State A has many districts that have a considerable range in various demographic characteristics and enrollment sizes. For the purpose of this study, we focused on districts that had more than 20 ELL students enrolled. This selection rule resulted in 38 districts in our study sample, which is only a fraction of all districts in State A. However, these districts tend to be larger districts. Overall, this sample retained 82% of the entire State A population and 94% of its

ELL population. Thus, sample reduction helps our study focus on issues around ELL students more clearly with no or trivial cost in terms of generalization. Table 1 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 state student sample, which confirms again that the study sample is representative of the states' entire population in the cohort.

Table 1

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%

Qualitative data. From January to July 2011, we conducted semi-structured interviews targeting the directors and coordinators of ELL programs in the 38 districts in our study sample. Participants were recruited after obtaining applicable approvals from university IRB and state offices. State A's Department of Education provided assistance in recruiting district directors and coordinators and recommended district personnel to interview. After an email invitation, we followed up with phone calls and additional email messages. Monetary compensation was not given to the participants, and all participation was strictly voluntary.

A total of 19 district personnel in charge of ELL programs participated in this study (participation rate = 50%). Although the low participation rate decreases the extent to which study findings can be generalized, the student characteristics and achievement levels of the 19 districts were comparable to those of the 38 districts, which was the original population that was targeted. From our sample of 19 participants, we interviewed five directors of multiple student services, five coordinators of ELL programs, and four directors of ELL programs. We also interviewed two coordinators of multiple student services and one of each of the following: an

ELL coach, an assistant director, and an assistant superintendent. Our participants ranged from less than one year to 11 years of experience and had an average of four years of experience directing ELL programs.

A major focus of the interview protocol concerned the standard criteria the district used to redesignate ELL students. Specifically, we asked districts for the criteria they used for different grade bands, that is, K-2 (primary), 3-5 (elementary), 6-8 (middle school), and 9-12 (high school). We also asked whether exceptions were made to their standard criteria. In other words, if a student did not meet the minimum requirements for redesignation established by the district (i.e., its standard criteria), would that student still be redesignated? If so, what evidence would they use to redesignate this student? Lastly, we asked who was responsible for the designation decision (and ultimately the criteria used for redesignation)—the district or the school. Interviews were conducted one-on-one over the phone and took approximately 30 minutes. All interviews were audio-recorded and later transcribed.

Analysis Methods

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 just 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.

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.

Specifically, we 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 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).

Model 1 shown below 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 time variable with values of -2, -1, and 0, 0, 0, 0 respectively for Grades 3, 4, 5, 6, 7, and 8, while the "Middle_Grade" variable is coded as the other time variable with values of 0, 0, 0, 1, 2, 3. 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

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

$$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Exit4})_i + \beta_{02}(\text{Exit5})_i + \beta_{03}(\text{Exit6})_i + \beta_{04}(\text{OtherELL})_i + r_{0i}$$
$$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Exit4})_i + \beta_{12}(\text{Exit5})_i + \beta_{13}(\text{Exit6})_i + \beta_{14}(\text{OtherELL})_i + r_{1i}$$
$$\pi_{2i} = \beta_{20} + \beta_{21}(\text{Exit4})_i + \beta_{22}(\text{Exit5})_i + \beta_{23}(\text{Exit6})_i + \beta_{24}(\text{OtherELL})_i + \beta_{25}(\pi_{0i} - \beta_{00}) + r_{2i} \quad 1(b)$$

Model 1 - Math

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

$$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Exit4})_i + \beta_{02}(\text{Exit5})_i + \beta_{03}(\text{Exit6})_i + \beta_{04}(\text{OtherELL})_i + r_{0i}$$
$$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Exit4})_i + \beta_{12}(\text{Exit5})_i + \beta_{13}(\text{Exit6})_i + \beta_{14}(\text{OtherELL})_i + \beta_{15}(\pi_{0i} - \beta_{00}) + r_{1i} \quad 2(b)$$

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.

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 the average growth and explores differences in post-reclassification growth across individuals and districts. We incorporate

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.

Similar specifications are used for these analyses that test correlates of more rapid growth rates over post-reclassification grades, with the subsample of reclassified ELLs only. With the subsample of reclassified ELLs, the same within-individual model as Model 1 is used. In order to examine student characteristics, or reclassification criteria that may be associated with subsequent success, the between-individual model is specified as a function of student demographics, grades at which students are reclassified, and student performance in components of EL reclassification standards just before reclassification. Data from the semi-structured interview concerning district reclassification practices consists of only 19 districts. Similar specifications were used in the analysis but we used a three-level multilevel model that adds a level of nesting clusters, districts, in a multilevel modeling framework.

Findings

Estimated Average Growth of Reclassified ELLs Compared to Other Students

Table 2 presents the results. Reclassified ELL students 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. In reading, ELL students reclassified at Grade 4 grow more rapidly on average by 1.1 points annually; those reclassified at Grade 5 on average by 0.9 points annually; and those reclassified at Grade 6 by 2.1 points annually. In math, ELL students reclassified at Grade 4, 5, and 6 grow more rapidly on average by about 1.0 points annually. All these estimates were statistically significant. 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.

Although this might indicate that existing reclassification decisions are, on average, supportive of ELL students' subsequent learning, such a conclusion should be tempered by two other findings from the study. First, when the estimated trajectories were superimposed on the achievement level bands designated by the state's standard (see Figure 1 for reading; and Figure 2 for math), there was no catch up of reclassified ELLs with their non-ELLs in an absolute sense. For example, students who exited at Grade 5 barely achieved the Proficient category in Grade 3,

and in Grade 4, and were in the Partially Proficient category. These students on average still barely achieved proficiency from Grades 6 to 8 after their reclassification.

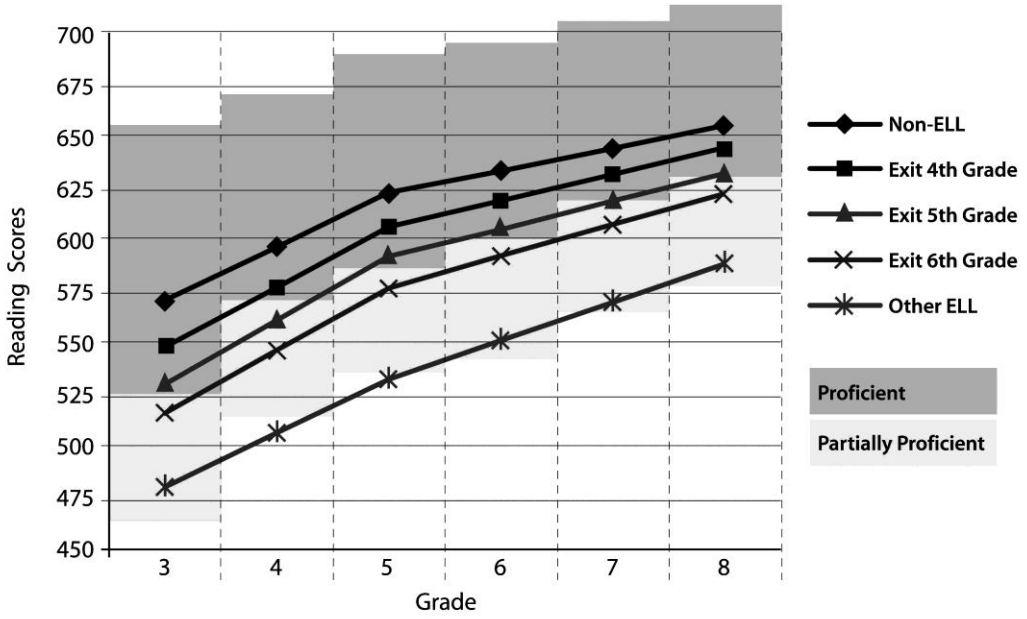

Figure 1. Estimated reading growth trajectories by ELL status and reclassified ELL status.

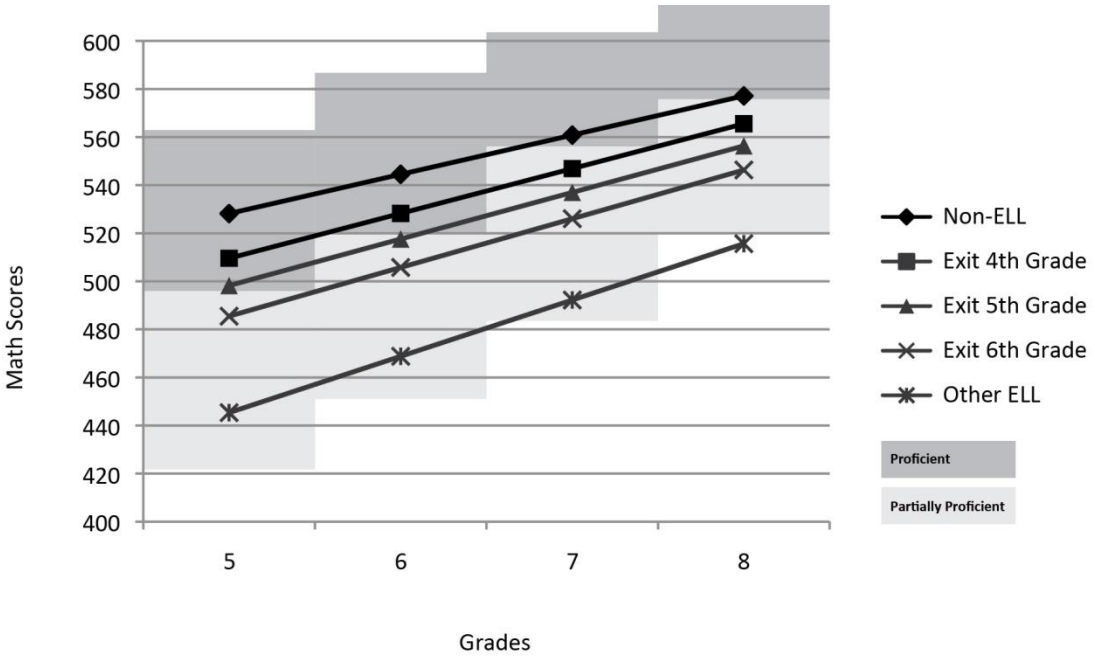

Figure 2. Estimated math growth trajectories by ELL status and reclassified ELL status
Note: The upper band of the shows the Proficient category, while the lower band shows the Partially Proficient category. Any scores higher than the upper band is the Advanced category; likewise, any scores lower than 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.

Secondly, the growth trajectories within groups were very heterogeneous both in reading and math (see Table 2; random effect estimates), which means that the estimated average trajectories might not carry much information for all reclassified ELLs. For example, 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.

Table 2

Results from Model 1-Reading

	Reading			Math		
Fixed effects	Coefficient	*SE*	*p*-value	Coefficient	*SE*	*p*-value
Model for status at Grade 5						
Intercept	622.12	0.33	0.000	528.18	0.40	0.000
EXIT4	-16.47	2.04	0.000	-18.61	2.42	0.000
EXIT5	-29.85	2.17	0.000	-29.96	2.55	0.000
EXIT6	-45.94	2.12	0.000	-42.70	2.48	0.000
OTHEREL	-89.92	1.04	0.000	-82.83	1.26	0.000
Model for growth rate during Elementary (Grades 3-5)						
Intercept	26.12	0.12	0.000	-	-	-
EXIT4	2.92	0.76	0.000	-	-	-
EXIT5	5.02	0.80	0.000	-	-	-
EXIT6	4.25	0.79	0.000	-	-	-
OTHEREL	-0.35	0.40	0.391	-	-	-
Model for growth rate during Middle, Post-reclassification (Grades 6-8)						
Intercept	10.65	0.07	0.000	16.32	0.08	0.
EXIT4	1.05	0.39	0.007	1.07	0.39	0.014
EXIT5	0.92	0.41	0.025	1.02	0.41	0.009
EXIT6	2.08	0.40	0.000	1.05	0.40	0.000
OTHEREL	2.93	0.23	0.000	1.48	0.23	0.000
Status at Grade 5	-0.56	0.01	0.000	-0.68	0.01	0.007

Random effects	Variance	*SE*	*p*-value	Variance	*SE*	*p*-value
Level-1 variance, temporal (within student)						
Grade 3	1532.80	17.30	0.000	1532.80	17.30	0.000
Grade 4	688.00	6.90	0.000	688.00	6.90	0.000
Grade 5	633.30	7.90	0.000	633.30	7.90	0.000
Grade 6	822.70	7.10	0.000	822.70	7.10	0.000
Grade 7	584.60	5.60	0.000	584.60	5.60	0.000
Grade 8	364.30	7.00	0.000	364.30	7.00	0.000
Level-2 variance (between student)						
Status at Grade 5	3557.40	27.80	0.000	3557.40	27.80	0.000
Growth rate Elementary	89.60	5.00	0.000	89.60	5.00	0.000
Growth rate Middle	50.30	1.30	0.000	50.30	1.30	0.000

Student Correlates of Post-Reclassification Growth

Table 3 presents the results. Demographic variables that are available from state data sets—ethnicity, grade levels at reclassification, and free or reduced lunch status – are included as predictors of growth parameters. Both in reading and math, significant predictors emerged as expected in terms of status, but only the ethnicity category turned out as a significant predictor of growth rates. Hispanic reclassified students on average grow significantly slower than other reclassified students in reading (estimate =-2.35, p-value =0.01).

Variables capturing reclassification criteria are also included as predictors of growth parameters. For example, we tested whether exiting with an ELP level of 5 is related to post-reclassification growth. Students exiting with the highest ELP level, as suggested in the state guidelines, tend to grow significantly more rapidly before reclassification (in reading). They also tend to exit at appreciably higher levels both in reading and math. The results diverge between reading and math for the growth rates after reclassification. 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 = -0.66, p-value = 0.50). 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 = -2.60, p-value = 0.02).

We tested other variables related to reclassification criteria such as exiting with different levels in state reading assessment, which comprises another main part of state guidelines on the reclassification of ELL students. None of these other variables were significant predictors of the growth rates after reclassification.

Table 3

Results from Model 2

	Reading			Math		
Fixed effects	Coefficient	*SE*	*p*-value	Coefficient	*SE*	*p*-value
Model for status at Grade 5						
Intercept	618.85	4.46	0.000	528.31	5.44	0.000
NATIVE	-29.31	10.66	0.006	-49.37	13.04	0.000
ASIAN	4.93	5.23	0.346	16.92	6.40	0.008
BLACK	-17.30	9.50	0.069	-18.21	11.61	0.117
HISPANIC	-17.87	4.35	0.000	-24.64	5.33	0.000
EXIT4	11.92	2.36	0.000	9.21	2.88	0.010
EXIT6	-15.75	2.41	0.000	-12.45	2.91	0.000
LOWSES	-14.49	2.44	0.000	-11.80	2.96	0.000
Model for growth rate during elementary (Grades 3-5)						
Intercept	31.44	1.95	0.000	-	-	-
NATIVE	-2.92	4.66	0.531	-	-	-
ASIAN	7.41	2.30	0.001	-	-	-
BLACK	-0.69	4.13	0.867	-	-	-
HISPANIC	-0.04	1.91	0.984	-	-	-
EXIT4	-2.40	1.03	0.020	-	-	-
EXIT6	-0.85	1.05	0.414	-	-	-
LOWSES	-1.06	1.06	0.315	-	-	-

Fixed effects	Reading			Math		
	Coefficient	*SE*	*p*-value	Coefficient	*SE*	*p*-value
Model for growth rate during middle, post-reclassification (Grades 6-8)						
Intercept	13.23	1.06	0.000	18.84	1.14	0.000
NATIVE	-1.38	2.34	0.555	-9.60	2.57	0.000
ASIAN	-0.59	1.08	0.586	0.96	1.18	0.418
BLACK	-0.10	1.98	0.962	2.05	2.16	0.343
HISPANIC	-2.35	0.91	0.010	-1.69	0.99	0.089
EXIT4	0.48	0.49	0.329	-0.29	0.53	0.586
EXIT6	0.79	0.50	0.111	-0.21	0.54	0.692
LOWSES	-0.27	0.50	0.597	-0.05	0.55	0.934
Status at Grade 5	-0.85	0.05	0.000	-0.67	0.05	0.000

District Reclassification Practices and Policies Associated with Post-Reclassification Growth

Data drawn from the interviews with district ELL coordinators provided information about district policies and practices regarding ELL reclassification. Results revealed that district varied considerably in the guidance and oversight they provided schools in reclassification decision-making. Some districts established standard criteria that should guide decisions; in others the criteria and decision-making were more collaboratively developed; and in still others, criteria and decision making were left completely at the school level. Districts also differed with regard to the range of criteria they encourages and with regard to how rigidly they adhered to established criteria.

Across these areas, our analysis suggested district differences in the evidence base and flexibility brought to bear in redesignation decisions. We hypothesized that district practices that may help student succeed after reclassification are those that guide schools to reclassify students based on multiple criteria, and carefully triangulate a range of evidence, including professional judgment, rather than relying on a narrow, uniform set of criteria. To create such a variable, we combined responses from various items. Specifically, we created binary indicators of the following practices, and summed them up to create a variable that reflects the extent to which districts use comprehensive, evidence-based practices for redesignation decisions:

- not requiring the highest level of ELD assessment (coded 1; otherwise 0)

- making exceptions from state guidelines (coded 1; otherwise 0)
- using additional criteria
- using course grades as part of additional criteria (coded 1;otherwise 0)
- either districts making reclassification decisions , or districts collaborate with schools with regard to reclassification decisions (coded 1); all decisions are delegated to schools (coded 0)

The 19 participating districts showed a range of values in this variable.

The district comprehensive, evidence-based practice variable is included as a predictor of growth parameters. Our preliminary analysis indicate that in districts exiting ELL students based on higher scores in the practice variable ELL students tend to grow significantly more rapidly subsequent to reclassification. The results were consistent between reading and math for the growth rates after reclassification. Students in districts with higher levels of the hypothesized practices of interest tend to grow at a significantly more rapid rate in reading as compared to students in other districts, holding constant performance status just before reclassification. Similar pattern of findings emerges in math. These findings are preliminary and we still plan to check whether the findings are robust to various specifications of models including covariate adjustment. Also, note that this set of analyses is intended to be exploratory and we did not adjust for multiple comparisons given the district sample size of 19.

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 also 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. 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 performance trajectories are compared to state-designated academic proficiency levels, there is little evidence that ELLs are catching up over the grades relative to their proficiency classifications: 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 scale values over the grades and insufficient to attain proficiency. Secondly, there is no evidence of former ELLs falling behind in academic growth after reclassification, either relative to their non-ELL peers or in terms of absolute academic proficiency levels. These findings 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 also provide positive empirical evidence to the validity of the existing reclassification system in the framework of this study.

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 Interrupted time series (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.

The grades at which reclassified ELL students exit are predictive of performance level. Gaps in performance status/level between EL and non-ELL students increase significantly from students reclassified at Grades 4 to those at Grade 5 and also from students reclassified at Grades 5 to those at Grade 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 Grade 4 may have received similar instruction and stayed in similar class settings to students reclassified in the Grade 5.

Lastly, this paper examined components of existing reclassification criteria and decision making processes by examining the relationship of ELP levels upon exiting to post-reclassification growth. In our framework, Reclassification criteria are valid as a set if they efficiently indicate readiness for mainstream classrooms, as indicated by success subsequent to

reclassification. 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.

This set of stringency analyses used the sample of ELL students reclassified around 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 learning 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.

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. The phase 2 of this study aimed to see whether such findings about reclassification criteria hold in light of specific district policies and practices. With the great heterogeneity in post-reclassification growth rates that were unexplained, we sought district-level variables such as reclassification criteria that might have explanatory value. As hypothesized, the preliminary results suggest that comprehensive, evidence-based district redesignation practices are associated with greater subsequent success in ELLs learning.

This finding adds potentially important evidence for creating ELL reclassification policies and practices. Although state A is a local-control state, many states reinforce uniform standards. based mostly on ELD scores and state reading assessments. Study findings about district policies and practices may suggest a need to adapt such uniform standards by making other evidence available and by combining information across sources in way that would not keep ELLs from exiting in the presence of a preponderance of evidence suggesting subsequent success. This finding may apply only to the grades of interest in this study, i.e., when students are about to move to secondary schools and face more challenging materials from a number of content areas. Study findings also may raise questions for further studies 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 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? More studies are warranted to obtain more concrete rules about optimal reclassification criteria.

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, NY: 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, NY: 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).

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).