LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

English Language Learners and Academic Achievement: Revisiting the Threshold Hypothesis

Yuliya Ardasheva
University of Louisville

Thomas R. Tretter
University of Louisville

Marti Kinny
Jefferson County Public Schools

This nonexperimental study explored the predictive strength of English proficiency levels on academic achievement of middle school students in a sample of 17,470 native English-speaking (NES) students, 558 English language learners (current ELLs), and 500 redesignated fluent English proficient students (former ELLs). Results of multilevel analyses indicated that after controlling for relevant student- and school-level characteristics, former ELLs significantly outperformed current ELL and NES students in reading (effect sizes: 1.07 and 0.52) and mathematics (effect sizes: 0.86 and 0.42). The results support Cummins's (1979, 2000) lower level threshold hypothesis predicting that upon reaching adequate proficiency in the language of schooling and testing, ELLs would no longer experience academic disadvantages. Refinements for the theory and directions for future research are discussed.

Keywords threshold hypothesis; bilingualism; ELLs' academic achievement; hierarchical linear modeling; achievement gap; academic achievement; language proficiency; redesignated fluent English proficient students; ELL reclassification/redesignation

Issues regarding the education of English language learners (ELLs) are gaining prominence on the national educational agenda (e.g., August & Shanahan,

---

The authors would like to thank Lourdes Ortega, the anonymous *Language Learning* reviewers, and Jim Cummins for their insightful comments. Any shortcomings of the study are our sole responsibility.

Correspondence concerning this article should be addressed to Yuliya Ardasheva, Department of Teaching and Learning, College of Education and Human Development, University of Louisville, Louisville, KY 40292. Internet: yyarda01@louisville.edu; jearda@gmail.com

DOI: 10.1111/j.1467-9922.2011.00652.x

2006; Solórzano, 2008) as well as in other English-speaking countries (Kim & Jang, 2009). Overall student achievement in U.S. schools will increasingly depend on the academic achievement of ELLs (Lazarin, 2006), who are expected to make up 40% of the total student population by 2050 (Goldenberg, 2008). Yet, as indicated by test scores in reading and mathematics, the ELL academic underachievement in comparison to native English-speaking (NES) students remains a reality (Fry, 2008; NCPPHE, 2005) and is often the cause of negative stereotyping regarding ELLs' intellectual abilities (e.g., Datnow, Stringfield, & Castellano, 2005; Lambert, 1978; Vollmer, 2000).

These statistics, however, are limited in terms of providing a complete and accurate picture of ELLs' academic ability. First, standardized academic achievement tests may provide downwardly biased estimates of what ELLs actually know (Abedi, 2004, 2007; Koretz, 2008; Solórzano, 2008)—particularly for students with strong academic backgrounds in their first languages (L1s)—and for many (e.g., Abedi, 2007; Solórzano, 2008), the question of defining the level of second-language (L2) proficiency at which achievement tests provide valid information about ELLs' academic development remains open. Second, what we know about achievement of language minority students (native speakers of languages other than English) is primarily based on the outcomes of the most struggling ELL subgroup—namely, students with low L2 proficiency (current ELLs).

Although much research has described and investigated the causes of the ELL-to-NES achievement gap (e.g., Cummins, 1981; Datnow et al., 2005; Grubb, 2008; Thomas & Collier, 2002), remarkably few studies have focused on academic achievement of former ELLs. Former ELLs are students who—upon reaching an English proficiency benchmark on standardized tests—are exited from current ELL status and no longer receive language support services. Two studies (Kim & Herman, 2009; NYCDE, 2009) conducted in states and school districts with historically large ELL populations and a long history of servicing the learning needs of this student group found that former ELLs either outperformed or performed on par with NES students on a number of academic achievement measures. Questions remain how the academic achievement of former ELLs compares to that of NES and current ELL students in the same classrooms, particularly in states and school districts in which ELL representation had been historically low.

By definition, ELLs are bilinguals—"people who need and use two (or more) languages in their everyday life"—accordingly, bilingualism in this study was defined as involving "the regular use of two (or more) languages" (Kormi-Nouri et al., 2008, p. 94). Bilingualism is often associated in the cognitive literature with certain benefits and advantages (e.g., better functioning of abstract representation, attentional control, and problem solving; see Adesope, Lavin, Thompson, & Ungerleider, 2010; Mindt et al., 2008) and, yet, ELLs as a population are negatively characterized in the educational research literature by a well-documented achievement gap (e.g., Fry, 2007, 2008; NCPPHE, 2005). One bilingual theory that may explain this apparent contradiction and the differential success of current versus former ELLs is Cummins's (1979) threshold hypothesis (TH), which predicts that "those aspects of bilingualism which might positively influence cognitive growth are unlikely to come into effect until the child has attained a certain minimum or threshold level of competence in a second language," provided that the child displays similar competencies in his or her L1 (p. 229). According to Cummins's TH, students with low levels of L1 and L2 proficiency ("partial" bilinguals) are likely to have "impoverished" interaction with their educational environments—"both in terms of input and output" (p. 230)—and thus experience academic disadvantages in schools; bilinguals with sufficient competency in one of their languages[1] ("dominant" bilinguals) would experience no such disadvantages; and students fully proficient in both languages ("additive" or "balanced" bilinguals) would enjoy cognitive and academic advantages associated with bilingualism. In other words, Cummins specified two thresholds of bilingual proficiency—the lower and the higher—and argued that students attaining these two thresholds would avoid the disadvantages (the lower threshold) and enjoy the advantages (the higher threshold) associated with bilingualism. Of note here, is that both in early TH formulations and more explicitly in later theory refinements, Cummins (1979, 2000) argued that disadvantages attributed to bilingualism are not an effect of bilingualism per se but rather are "a result of discriminatory schooling [. . .] when schools deny bilingual students opportunities to access literacy and comprehensible academic language in both L1 and L2," thus denying students the benefits of additive bilingualism and often leading to students' progressively falling "further behind grade expectations in their functional command of academic registers" (2000, p. 100).

Much bilingual research, however, has been conducted in multilingual or two-languages-in-the-curriculum contexts and has focused, in essence, on testing the higher level TH by comparing low- versus high-level bilinguals (e.g., Andreou & Karapetsas, 2004; Lee & Schallert, 1997; Schoonen, Hulstijn, & Bossers, 1998) or "balanced" bilinguals versus monolinguals (e.g., Costa, Hernández, & Sebastián-Gallés, 2008; Kormi-Nouri et al., 2008) on a number of cognitive, metacognitive, and academic outcome measures. With some notable exceptions (see a body of work by Bialystok and colleagues conducted in

Canada; e.g., Bialystok, 1986, 1999; Bialystok, McBride-Chang, & Luk, 2005), research on the correlates of bilingualism with regard to ELLs—particularly in nonbilingual educational environments in the United States—remains limited (see Adesope et al., 2010). This may be due, in part, to the noted asymmetry in language skills development in ELLs schooled in nonbilingual educational environments (Proctor, August, Carlo, & Barr, 2010) and difficulties in measuring L1 proficiency in schools in which a great many L1s are spoken. At the same time, a number of researchers (e.g., Adesope et al., 2010; Lee & Schallert, 1997) have called for expanding research on the correlates of bilingualism to immigrant L2 learners for three reasons: (a) demographic changes and increased instances of bilingualism and multilingualism in the United States as well as other English-speaking countries (Adesope et al., 2010; Goldenberg, 2008; Kim & Jang, 2009); (b) a combination—characteristic of ELL learners—of typically lower levels of L1 literacy skills with advantages associated with contextual support from being immersed in L2-rich environments (Lee & Schallert, 1997); and (c) the assumption that cognitive benefits associated with bilingualism "may be of use to bilingual speakers in classrooms where the language of instruction is not their native language" (Adesope et al., 2010, p. 231).

The present study was undertaken with two objectives in mind. One objective was to extend research on academic achievement of former ELLs' to places with traditionally small ELL populations. Another objective was to test—in an exploratory fashion—the lower level TH by operationally defining lower level threshold L2 proficiency as the reclassification from current to former ELL status based on standardized English language proficiency measures and by comparing academic achievement of middle school ELLs at and below this threshold against that of monolingual, NES students after controlling for individual and school socioeconomic status (SES). Because of a lack of L1 standardized measures, this study was unable to measure L1 literacy skills. Instead, based on a premise that all healthy children develop adultlike grammar in their L1s without instruction, typically by the age of 5 (Chomsky, 1959; Fromkin, Rodman, & Hyams, 2007; MacSwan, 2000) and precedence in the literature (e.g., Bialystok, 1986, 1999; Kormi-Nouri et al., 2008), we accepted the use of a home language other than English as adequate evidence of middle school ELLs' L1 oral proficiency, defined as the ability to comprehend and use language in oral communication. Thus, based on the fact that both of our ELL groups displayed high levels of at least oral proficiency in English (see Appendix A) and on evidence of typically lower (below grade-level norms) levels of L1 literacy skills in nonbilingual programs (Proctor et al., 2010), we applied Cummins's terminology and defined the two groups—current and former ELLs—respectively as "partial" and at least "dominant" (English) bilinguals.

Before proceeding, we want to acknowledge that the comparison among current ELL, former ELL, and NES students may confound academic achievement with levels of English literacy skills. This comparison, however: (a) represents a common practice in educational research—at least with respect to the NES-to-ELL academic achievement gap literature (e.g., Fry, 2007, 2008; NCPPHE, 2005; Reardon & Galindo, 2009); (b) reflects federal policies requiring separate tracking of current and former ELLs' achievement (Lazarin, 2006); and (c) has an important advantage, as it allows controlling for the instability of the ELL population (i.e., a constant influx of low English proficiency students and a constant exit of students who have reached an English proficiency benchmark; Kim & Herman, 2009; NYCDE, 2009). With the present study, we seek to offer several important methodological and theoretical advancements in the TH research and to provide insights into the relationship between levels of L2 proficiency and academic achievement among ELLs, an important group of students in U.S. schools.

Study Background

Bilingualism as Advantage

Evidence from psychology and neurology has suggested that speaking more than one language has "profound impact on cognitive development, as well as brain structure and function" in bilingual students (Mindt et al., 2008, p. 261). Studies have documented balanced bilinguals outperforming matched monolingual controls on a variety of cognitive tasks (e.g., Bialystok, 1986, 1999; Bialystok & Martin, 2004; Costa et al., 2008; Kaushanskaya & Marian, 2009; Kormi-Nouri et al., 2008). In a recent meta-analysis of 63 studies involving different L1-L2 combinations, for example, Adesope et al. (2010) found reliable associations between bilingualism and several cognitive benefits in the areas of metalinguistic and metacognitive awareness, working memory, abstract and symbolic representation, attentional control, and problem solving. Although several explanations have been advanced in the literature (see a discussion in Kormi-Nouri et al., 2008), bilingual cognitive advantages have been most often attributed to the "added practice" in exercising *cognitive control*—that is, the ability to focus attention and resist interference (Bialystok, 1999; see Adesope et al., 2010; Costa et al., 2008; Mindt et al., 2008).

By definition, bilinguals need to coordinate two languages on an everyday basis, which often involves consciously focusing on the target language and inhibiting the nontarget language by applying "the same executive functions used generally to control attention and inhibition" (Green, 1998, as cited in Mindt et al., 2008, p. 268). This affords bilinguals extensive practice with selection and inhibition and may ultimately lead to greater efficiency in exercising cognitive control (Bialystok, 1999; Costa et al., 2008). Brain imagining research has provided some evidence in support of this hypothesis. Studies have found evidence of the brain's adaptation to the demands of a dual-language environment (e.g., gray-matter density increase and greater activation of brain regions associated with conflict resolution and higher order attentional processes, including selective attention and inhibition; see a review by Mindt et al., 2008). Further, there is some evidence suggesting earlier development of cognitive control functioning in young bilinguals (Bialystok, 1999).

Developing proficiency in an additional language, however, is a challenging process requiring gradual accumulation of skills over time. Evidence suggests that ELLs may require 2–5 years to develop mastery of English oral skills (e.g., sound discrimination, vocabulary, listening comprehension, oral expression; Hakuta, Butler, & Witt, 2000) as well as mastery of syntactic, morphological, and pragmatic skills (MacSwan & Pray, 2005), and up to 7 or more years to reach high levels of literacy skills comparable to average NES student performance on standardized tests in reading (Collier & Thomas, 1989; Thomas & Collier, 2002). Eventually, however, the returns on bilingualism may produce academic benefits for the learner (Cummins, 1981; Diaz & Klingler, 1992), particularly in bilingual programs in which students receive instruction through both the L1 and L2 (Rolstad, Mahoney, & Glass, 2005; Slavin & Cheung, 2005). Yet, as of now, "there is no cross-disciplinary consensus regarding the nature of language proficiency and its relationship to academic development" among ELLs (Cummins, 2008, p. 81).

The Threshold Hypothesis: Promises and Critiques

In putting forth TH, Cummins (1979, 2000) argued that levels of bilingualism have a mediating effect on cognitive and academic functioning of the students and proposed "not one, but two, thresholds"—namely, the lower and the higher level of bilingual competence (1979, p. 230). According to Cummins, low levels of L1 and L2 academic language proficiency[2] "[limit] children's ability to benefit cognitively and academically from interaction with their environment through those languages" (2000, p. 175). Such disadvantages, Cummins argued, are likely to be mitigated once bilinguals reach the lower level threshold—that is, once they develop substantial literacy skills in the language of instruction. The higher level threshold, in turn, specifies the "continued development of both languages into literate domains" as a necessary condition under which bilinguals will experience "enhanced cognitive, linguistics, and academic growth" (2000, p. 37). The TH offers two advantages to researchers and educators. First, the theory explains differential cognitive and academic outcomes among bilingual students noted in the literature. Second, the theory offers a potentially testable hypothesis predicting that academic and cognitive outcomes of ELLs—in comparison to those of monolingual, NES students—would be comparable at the lower level threshold and potentially superior at the higher level threshold.

Although Cummins's TH continues to generate interest in the research community (e.g., Andreou & Karapetsas, 2004; August & Shanahan, 2006; Lasagabaster, 1998; Schoonen et al., 1998), it has also been criticized on both methodological and theoretical grounds (MacSwan, 2000; see also Takakuwa, 2005). Among other methodological shortcomings, Takakuwa (2005) has criticized past research guided by the TH for defining the thresholds in a relative, rather than absolute, sense by setting up arbitrary, within-group thresholds (e.g., median split, mean split, years of L2 study) that are based on a variety of L1 and L2 proficiency measures (e.g., standardized tests, researcher-developed measures such as grammaticality judgment and cloze tests)—thus yielding "countless threshold levels" and rendering the hypothesis "meaningless" (p. 2230)—as well as for overreliance on statistical significance and for exploring primarily relationships of association rather than those of causality. Takakuwa argued that these shortcomings limited the generalizability and replicability of the TH research and called for more rigorous investigations of the theory.

From a theoretical perspective, MacSwan (2000) has criticized the TH, among other things, for not differentiating between oral language and literacy skills.[3] At the same time, two ways—linked to the latter oral-literacy skills distinction—in which bilingual experiences may positively contribute to student academic outcomes have been advanced in the literature by Bialystok et al. (2005). The first contribution pertains to what we call a *biliteracy advantage*, which obtains from transfer across two languages of linguistic and academic skills such as cognate knowledge or reading strategies (August & Shanahan, 2006; Cummins, 1981; Genesee, Lindholm-Leary, Saunders, & Christian 2005). The second contribution lies with what can be termed *cognitive processing benefits*, including better functioning of selective attention (Bialystok, 1986, 1999; Bialystok & Martin, 2004), conflict resolution (Costa et al, 2008), and memory (Kaushanskaya & Marian, 2009; Kormi-Nouri et al., 2008), often assessed by nonlinguistic measures such as reaction time or number of correct trials in response to visual stimuli (e.g., Bialystok, 1999; Costa et al., 2008). Empirical evidence has suggested that in order for knowledge and skills transfer to occur, bilinguals require substantial levels of literacy skills in both languages (Clarke, 1979, 1980; Lee & Schallert, 1997; Proctor et al., 2010). On the other hand, research on preschool bilinguals (e.g., Bialystok, 1986, 1999; Adesope et al., 2010) has suggested that cognitive processing benefits may be less dependent on biliteracy skills than the biliteracy advantage. In a meta-analysis of 12 studies conducted in preschool settings, Adesope et al. (2010) found that young bilinguals—whose literacy skills across languages are likely to be emergent, at best—outperformed (effect size = 0.63) monolinguals on measures of symbolic representation, attentional control, and problem solving. This suggests that differentiating between oral language and literacy skills may be a necessary step in advancing TH research and our understandings of the relationships between language proficiency and cognitive and academic development among bilinguals. We return to this point in the Discussion section.

ELL Academic Achievement

The current ELLs' achievement gap has been the topic of much research, publication, and discussion (e.g., NCPPHE, 2005; Reardon & Galindo, 2009). Analyses of the U.S. Department of Education databases for the 2004–2005 school year in five states with large ELL populations, for example, indicated that ELLs were less likely to score at proficiency levels in reading and mathematics than did NES students (Fry, 2007, 2008). These results, however, are not surprising, as the main criterion for being identified as current ELL is that the student is behind in his or her English skills development. On the positive side, a recent report from the New York City Department of Education (NYCDE, 2009), where ELLs constitute 26% (about 312,000 students) of the school population, found evidence of strong academic performance by former ELLs. Over a 5-year period (2003–2008), the former ELL group consistently produced the highest percentage of students scoring proficient in reading and mathematics in Grades 4 and 8 in comparison to current ELL and NES students. Current and former ELLs tended to "experience wider and steeper gains, trending more closely with each other as opposed to non-ELLs" (p. 24). In another study, Kim and Herman (2009) analyzed data from three states with large ELL populations and found that—depending on time since reclassification and stringency of reclassification criteria—elementary and middle school former ELLs outperformed or performed on par with NES students on standardized tests in reading, mathematics, and science. Regardless of the reclassification criteria stringency, students with 2 or more years since reclassification consistently outperformed all other comparison groups. These findings highlight English proficiency as an important predictor of ELLs' academic achievement.

English Proficiency and Academic Achievement

Although no universally accepted definition of English proficiency across disciplines (Cummins, 2008) or across states (Solórzano, 2008) exists, English proficiency—most often defined as language-specific knowledge (e.g., vocabulary, structures, contextually appropriate language use)—has been identified as a strong student-level predictor of academic achievement in ELLs (Ardasheva, 2010; Mahon, 2006; Solórzano, 2008; Suárez-Orozco, Suárez-Orozco, & Todorova, 2008; Yoko, 2007). Limiting effects of low levels of L2 proficiency on L2 academic outcomes have been long recognized (Clarke, 1979, 1980; Schoonen et al., 1998). Reading comprehension research found that low levels of L2 proficiency "short circuited" or limited the extent to which students were able to use their L1 academic skills—namely, cognitive skills such as reading strategies (Clarke, 1979, 1980; Schoonen et al., 1998) and metacognitive skills such as task-knowledge and metacognitive strategies (Schoonen et al. 1998)—to support their L2 performance. Lee and Schallert (1997) found that high school students' L2 proficiency accounted for a substantially larger proportion of variance in L2 reading performance than did L1 reading ability (56% and 30%, respectively),[4] and similar results were found for elementary ELLs by Mahon (2006). Likewise, Solórzano (2008) reported a significantly higher performance on native language standardized tests versus comparable English tests both in current and former ELL elementary students. The current No Child Left Behind law (NCLB, 2002) recognized that there are confounding effects of language and content knowledge that contribute to standardized test scores of ELLs and encouraged states to provide language accommodations when students are tested in English or, when feasible, to test ELLs' content knowledge in their native languages (Lazarin, 2006).

Programs, Accommodations, Services, and Support for ELLs

Research (e.g., Rolstad et al., 2005; Slavin & Cheung, 2005; see also August & Shanahan, 2006; Genesee et al., 2005) has indicated that bilingual education programs benefit ELLs' educational outcomes. According to Goldenberg (2008), a total of five independent meta-analyses confirmed these results. The effectiveness of bilingual education has been attributed to transfer of linguistic and academic skills across languages (e.g., August & Shanahan, 2006; Cummins, 1981; Genesee et al., 2005). For example, in their systematic review of over 200 articles and reports on educational outcomes of ELLs, Genesee et al. (2005) concluded that bilingual proficiency and biliteracy have a positive relationship with academic achievement in both English and native languages. The authors noted that positive correlations between L1 and L2 reading, L1 reading and L1 mathematics, and L2 reading and L2 mathematics suggest "complex but supportive interdependencies in the language, literacy, and academic development" of ELLs (p. 376). Other instructional supports identified as effective in advancing ELLs' language and academic outcomes include linguistic accommodations and explicit L2 instruction. In their randomized control trial study of about 1,000 students, Kopriva, Bauman, Cameron, and Triscari (2009) found that the significant NES-ELL achievement gap observed on traditional tests disappeared when students were tested using alternative, computed-based assessments with minimized language demands (i.e., introducing pictures, audio support in L1 or L2, animated directions for task completion). Further, the results of NES students on the traditional versus alternative tests were statistically the same, suggesting that linguistic accommodations did not compromise the validity of the alternative tests. Likewise, Abedi and Lord (2001) found that linguistic modifications (e.g., shortening nominal phrases, substituting passive voice and low-frequency vocabulary) of National Assessment of Educational Progress (NAEP) test items in mathematics resulted in slight, but significant improvement in performance of eighth-grade ELLs. Further, research (see reviews by DeKeyser, 2003; Norris & Ortega, 2000; Spada & Tomita, 2010) has suggested that L2 development benefits from varied forms of explicit language instruction (e.g., focus on form, enhanced input), including in content-area classrooms (Zwiers, 2006, 2007).

SES and Academic Achievement

Ample literature (e.g., Kao & Thompson, 2003; Thomas & Collier, 2002; Wilde, 2009a, 2009b) has suggested that individual students' SES, as measured by eligibility for free or reduced-price lunch or by parental level of education, has an impact on academic careers of both NES students and English learners. For instance, students with low SES were found to acquire English proficiency at a lower rate than more advantaged children (Carhill, Suárez-Orozco, & Páez, 2008; Hakuta et al., 2000; Páez, 2002). Using structural equation modeling techniques, Yoko (2007) found that SES, along with length of residence in the United States, native language, and migrant status accounted for 90.3% of the variance in reading achievement and for 58.7% of the variance in mathematics achievement in a sample of 1,985 sixth graders. The 2008 NAEP results for NES students showed that across Grades 4, 8, and 12, low-SES students had an average 22-point gap in reading and mathematics as compared to high-SES students; low-SES current ELL students had the lowest scores (Wilde, 2009a, 2009b). At the same time, Proctor et al. (2010) found that SES measures had no significant association with native language literacy skills.

School Effects on Academic Achievement

School effects on academic achievement of both NES students and ELLs have long been recognized (Fry, 2008; Grubb, 2008; Kao & Thompson, 2003). Suárez-Orozco et al. (2008) found that school poverty rate (percentage of low-SES students in a school) as well as racial representation, the percent of students scoring proficient on the state English Language Arts test, and the average attendance rate were significant predictors of ELLs' test scores in reading and mathematics; the four-predictor model accounted for about 32% of the variance in the test scores. Further, attending schools with higher SES and achievement profiles appears to have a positive impact on student English language development (Carhill et al., 2008; Hakuta et al., 2000). Hakuta et al. (2000) estimated that students from high-poverty schools were 1 year behind in English acquisition compared to students from more economically advantaged schools. Using a two-level hierarchical linear model, Yoko (2007) found that, among school-level variables, school poverty rate had the most influence on academic achievement in reading, writing, and mathematics among fourth- and sixth-grade students, predicting lower performance among ELLs attending higher-poverty schools. Yet, a trend common to ELL populations is to attend schools with "high levels of students living in or near poverty" (Fry, 2008, p. i).

In sum, previous studies suggest a hierarchical relationship between student- and school-level characteristics as they relate to academic outcomes.

Present Study

The present study employed a nonexperimental design and analyzed cross-sectional data collected by a large Midwestern school district in the 2007–2008 school year in order to investigate the following two research questions:

1. How does academic achievement in reading of middle school former ELLs compare to that of NES and current ELL students controlling for student SES and school poverty rate?
2. How does academic achievement in mathematics of middle school former ELLs compare to that of NES and ELL students controlling for student SES and school poverty rate?

The analytical procedures included descriptive statistics and hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002). A two-level hierarchical linear model was deemed appropriate because, as suggested by the literature, student academic achievement may be differentially impacted by different English proficiency levels and individual and school SES characteristics.

Method

Population and Site

The targeted population for this study were middle school redesignated fluent English proficient (former ELL) students; low English proficiency (current ELL) and native English-speaking (NES) students served as comparison groups. Middle school students were chosen because this age bracket was more likely to include sufficient numbers of current and former ELLs.

In the 2007–2008 school year, the district in which this study was conducted serviced over 98,000 students; ELLs comprised about 5% of the student body. Annual Measurable Objectives in reading and mathematics (i.e., percent of students scoring proficient or distinguished on state standardized tests) were met by the total student body, but not by current ELLs in reading (KDE, 2009). The district used the results of two measures—Home Language Survey (a survey of the student and family home language profile) and a state-mandated English language proficiency (ELP) placement test—to identify current ELLs and to screen for candidates for English-as-a-second-language (ESL) services (families had an option of waiving these services). Content-based instruction (integrated language and content instruction) served as the primary ESL service supported by bilingual aides, mostly Spanish-speaking.

A student's continued status as a current ELL was determined through annual state-mandated ELP tests (see below). Although some school districts use a combination of criteria for exiting students from current ELL status (e.g., academic achievement in core content areas plus ELP data; Solórzano, 2008), the district in this study made reclassification decisions based on measures of English proficiency only, not on academic content measures. Once students scored proficient on ELP tests, they were reclassified as former ELLs, were exited from ESL services, and entered a "monitoring" phase for a minimum of 2 years. Monitoring served to ensure that former ELLs were able to (a) meaningfully participate in all-English classrooms without the use of simplified English materials, (b) perform on grade level, and (c) have access to language support instruction services if needed (KDE, 2008).

Participants

The study employed purposive sampling at the school level. To be included in the study, a school had to (a) be a regular education school and (b) have a population of current and former ELLs within its body. Out of 24 middle schools in the district, 22 schools met these criteria. Eight of these schools provided ESL services; 11 schools were Title I (i.e., schools classified as low income and receiving additional federal funds). Five middle schools both provided ESL services and were Title I schools. Data of all Grade 6, 7, and 8 current ELLs, former ELLs, and NES students enrolled in the selected schools were included in the analyses, after excluding Initially Fully English Proficient students (IFEP; students whose home language is other than English, but whose high ELP scores prevented them from ever being classified as current ELLs; $N = 142$).[5] Data across the three grade levels were aggregated to form a single middle grades sample. This decision was made because some of the schools lacked either current or former ELLs in certain grade levels, which would have led to the exclusion of these schools from data analyses. The total student sample included 18,523 students (17,470 NES, 558 current ELLs, 500 former ELLs); average, within-school sample sizes for current and former ELLs were 25 (range: 1–108) and 23 (range: 2–60), respectively.[6] The average age of students was 13.63. Fifty-seven percent were eligible for free or reduced-price lunch and 49% were females. Current and former ELLs spoke 48 home languages. The top four languages included Spanish (47.2%), Bosnian (10.1%), Vietnamese (5.3%), and Maymay (4.7%). Descriptive statistics by language groups are summarized in Appendix B.

Analyses using χ^2 for categorical variables and ANOVA and t-tests for continuous variables indicated that the three comparison groups differed in their demographics. There were fewer females in the current ELL group than in the former ELL and NES groups, $\chi^2(2) = 11.97, p < .01$ (see Appendix B). More current and former ELLs were eligible for free and reduced-price lunch than were their NES counterparts, $\chi^2(2) = 359.90$, $p < .001$. Chi-square analysis also revealed differences in racial and ethnic composition, $\chi^2(10) = 6{,}693.62$, $p < .001$. Whereas NES were primarily White and Black, Hispanic students constituted a majority in current and former ELL groups. Finally, an ANOVA analysis revealed significant age differences, $F(2, 18{,}520) = 41.28, p < .001$. Tukey post hoc tests indicated that former ELLs were significantly younger than NES and current ELL students and that NES students were younger than current ELL students. Average time in the U.S. schools (years since first U.S. school enrollment starting in kindergarten) was 6.63 ($SD = 1.84$) and 4.72 ($SD = 2.37$) years for former and current ELLs, respectively. This difference was statistically significant, $t(1027) = -14.191, p < .001$, indicating that, on average, former ELLs had been in the U. S. school system longer than current ELLs by about 2 years. (The majority of current ELLs—63%, in comparison to 24% of former ELLs—had been in the U.S. schools for about 5 or less years.) The HLM analyses addressed group demographic differences as described in subsequent sections.

Measures

Academic Achievement

Student progress in meeting state expectations across individual content areas and grade levels is measured in this district with the criterion-referenced Kentucky Core Content Tests (KCCT; see KDE, 2005).[7] KCCT tests are aligned with the state's core content for assessment, which, in turn, define the content and scope of the curriculum provided by the district. Psychometric properties of the tests are evaluated and reported yearly on the state's department of education Web site. Raw scores are converted into scale scores (range: 0–80); item response theory (IRT) scaling allows for comparison of scale scores across grade levels. Based on the scale scores, students are assigned to one of four proficiency levels (range: Apprentice to Distinguished). Each proficiency level spans approximately 20 scale score points across the entire 0–80 range. These proficiency levels are a core component of the basis on which school overall performance is judged. Current ELL students are tested with accommodations (e.g., reading to the student, paraphrasing directions, oral word-for-word translation, extended time). In the present study, reading and mathematics achievement on the KCCT was inspected.

Reading Achievement

The reading test assesses students' foundational skills (e.g., suffixes, prefixes, synonyms), developing understanding (e.g., making summaries, predictions), text interpretation skills (e.g., identifying authors' opinions), text response skills (e.g., connecting text to real world issues), and the ability to demonstrate a critical stance. Appendix C displays the distribution of KCCT proficiency levels in reading disaggregated by language proficiency groups.

Mathematics Achievement

The mathematics test assesses students' knowledge in five domains: number properties and operations (e.g., ratios, proportions); measurement (e.g., measuring physical attributes); geometry (e.g., shapes, coordinate geometry); data analysis and probability (e.g., data representations, probability); and algebraic thinking (e.g., equations, inequalities). Appendix C displays the distribution of KCCT proficiency levels in mathematics disaggregated by language proficiency groups.

English Proficiency

Prior to 2007, the district used Language Assessment Scales–Reading/Writing and –Oral (LAS-R/W and LAS-O, respectively; Duncan & De Avila, 1988, 1994). LAS-R/W measured English reading and writing skills (vocabulary, fluency, reading comprehension, mechanics, and usage); proficiency levels ranged from 1 = Non-Reader/Writer to 3 = Competent Reader/Writer. LAS-O measured English speaking and listening skills (phonology, vocabulary, syntax, and pragmatics); proficiency levels ranged from 1 = Non-English Speaker to 4/5 = Fluent English Speaker. Information on LAS validation is limited. Two studies (Dalton, 1979; Schrank, Fletcher, & Alvarado, 1996) reported high correlations between LAS-O scores and teachers' ratings of student English proficiency (.73 and .76, respectively). The test was widely criticized for its low predictive validity of ELLs' academic achievement (see a discussion in Solórzano, 2008). Aware of this shortcoming, the district used information from LAS scores—a composite score of 10 or higher, with the most common proficiency level profile of 3–2–5 (Reading-Writing-Speaking)—in combination with ESL teacher recommendation to reclassify students as former ELLs. A total of 336 former ELLs in this sample were reclassified based on LAS.

Beginning in 2006–2007, the state adopted Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs®: Form 201; WIDA, 2008). The instrument—henceforth referred to as ACCESS—is composed of four subsets measuring English proficiency in four language domains: listening, reading, writing, and speaking (reliability indexes range: .82–.97 [Kenyon, 2006]; proficiency levels range: 1 = Entering to 6 = Reaching). In order to gauge students' development of the proficiencies needed to deal with the language of the school, the test is aligned with the state content area standards in reading, mathematics, and science (standard-to-standard reliability indexes range: .82–.98; Cook & Wilmes, 2007). Student responses are evaluated against three criteria measuring language-specific knowledge: (a) *complexity*, the amount and quality of the speech; (b) *vocabulary usage*, the specificity of words or phrases; and (c) *language control*, control over mechanics, syntax, and semantics (WIDA, 2008). A composite score of 5 indicates a judgment that a student is ready to perform in English-only classrooms without additional language support and serves as a criterion for reclassification into former ELL status. In this sample, 162 former ELLs were reclassified based on ACCESS; two students had missing test-of-reclassification (LAS vs. ACCESS) data.

Appendix A displays the distribution of ACCESS proficiency levels for current and former ELLs. LAS proficiency level data for students reclassified prior to 2007 were not available; this was not of concern because the HLM analyses reported in this study were based on reclassification status operationalized as a categorical variable rather than on specific English proficiency scores.

Study Variables

Study variables were grouped into two categories, student level and school level. Student-level variables included gender, age, three levels of English proficiency (NES [baseline], current ELL, former ELL), and individual SES. Gender and age were introduced into analyses to control for the between-group demographic differences noted earlier. Although the comparison groups also differed in racial composition, we chose not to control for race. As noted by Koretz (2008), combining diverse groups under the same label (e.g., including newly arrived—thus, often low income—and well-established Asian American families under Asian/Pacific Islander category) lends poorly for the interpretation of any outcome comparisons by masking within-group differences. In our judgment, SES and English proficiency status better represent differences in sociocultural experiences when nonimmigrant and immigrant students are compared.

School poverty rate (the percent of students eligible for free or reduced-price lunch in a given school) served as a school-level variable. Originally, the authors considered one additional school level variable–namely, Accountability Index (a state standardized measure of aggregated average academic and nonacademic—such as attendance and retention—school performance

Table 1 Description of the study variables

Variables	Descriptions
Dependent variables	
Reading achievement (READ)	KCCT scale score (0–80) in reading in 2008
Mathematics achievement (MATH)	KCCT scale score (0–80) in mathematics in 2008
Student-level predictors	
Gender	0 = female; 1 = male
Age	In years
Current ELL	0 = non-current ELL; 1 = current ELL
Former ELL	0 = non-former ELL; 1 = former ELL
Native English-speakers (NES)	0's on current and former ELL status (reference group)
SES	0 = paid lunch; 1 = free or reduced-price lunch
School-level predictor	
School poverty rate (POVERTYR)	% of students qualified for free or reduced-price lunch

measures). However, preliminary analyses indicated that school poverty rate shared 84% of the variance with the Accountability Index ($r = -.92$, $p < .01$, $R^2 = .84$). In order to avoid multicollinearity and because of the small number of second-level units (22 schools), only the school poverty rate was retained for further analyses. Table 1 provides the coding scheme for each of the variables.

Data Collection Procedures

KCCT, LAS, and ACCESS were administered under secure conditions following the timetable, regulations, and procedures set by the district. English proficiency tests were administered by ESL teachers in January/February of each year. LAS was scored by ESL teachers using a scoring rubric provided by the test developer. ACCESS was scored by ESL teachers (speaking component) and by the test developer (reading, listening, and writing components). KCCT tests were administered during the spring testing window (2 weeks in May) by content-area teachers and scored by a third party.

Of the over 18,000 students in the middle school sample, a total of 276 students (less than 2%) were missing KCCT scores. This group included 19 current ELLs, 1 former ELL, and 256 NES students. Among these students, 210 pupils received special education services (9 current ELLs; 210 NES). All cases with missing test scores were omitted from the analyses. No other student- or school-level data included in HLM analyses were missing.

Data Analyses

In order to compute the unique contributions and interaction effects of student- and school-level variables on academic achievement, the present study estimated a two-level hierarchical model using maximum-likelihood estimation. Student-level variables included age, gender, current ELL, former ELL, and SES. Because we used age and gender to control for group differences identified earlier and assuming that their effects did not vary across schools, these two variables entered the equation as fixed. The remaining variables of substantive interests were specified as random. All categorical variables at the student level were dummy coded (0, 1) and entered the equation uncentered. Age, a continuous variable, was group mean centered.[8] Poverty rate, a school-level variable, was grand mean centered. The Level 1 equation is specified as

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{GENDER}_{ij}) + \beta_{2j}(\text{AGE}_{ij}) + \beta_{3j}(\text{current ELL}_{ij})$$
$$+ \beta_{4j}(\text{former ELL}_{ij}) + \beta_{5j}(\text{SES}) + r_{ij},$$

where Y_{ij} is the test score for student i in school j (Y = reading OR mathematics), β_{0j} is a mean KCCT score in reading or mathematics for the reference student (female, average age, native English-speaker, high-SES student in a school j that is average on poverty rate; see Table 1 for the coding scheme that established the reference group), β_{pj} is the expected change in Y_{ij} for a unit change in X (X = GENDER, AGE, current ELL, former ELL, SES), and r_{ij} is the unique random effect for student i in school j on the test score (Y_{ij}). The full Level 2 model for estimating the means was

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{POVERTYR}) + u_{0j},$$

where γ_{00} is the mean KCCT score in reading or mathematics for a school j that is average on poverty rate, γ_{01} is the change in β_{0j} for a 1% change in school poverty rate (POVERTYR), and u_{0j} is the residual variance of school j after controlling for POVERTYR. The full Level 2 model for estimating the Level 1 slopes was

$$\beta_{pj} = \gamma_{p0} \qquad \text{for } p = 1, 2 \text{ (to control for GENDER and AGE)},$$

$$\beta_{pj} = \gamma_{p0} + \gamma_{p1}(\text{POVERTYR}) + \mathrm{u}_{pj} \qquad \text{for} p = 3, 4, 5,$$

where p = 1–5 are coefficients from Level 1 model, γ_{p0} is the predicted (average) X-test score slope[9] (X = GENDER, AGE, current ELL, former ELL, SES) in reading or mathematics across schools, γ_{p1} is the change in the predicted X-test score slope (β_{pj}) for a 1% change in POVERTYR, and u_{pj} is the residual variance associated with school j on the predicted X-test score slope (β_{pj}) after controlling for POVERTYR.

Results

Descriptive Statistics

For the combined sample of 18,523 students, the overall mean scale scores were 44.6 (SD = 18.6) and 39.6 (SD = 22.4) for reading and mathematics, respectively. At the school level, for the sample of 22 schools the mean poverty rate was 56.2% (SD = 20.6) of students on free/reduced-price lunch status. The range of poverty rates for these schools was 19–89%. In the following subsections, HLM results are presented for both reading and mathematics. For ease of inspection, results for reading achievement are provided first, always immediately followed by results for mathematics achievement in brackets.

Model Specifications

In order to estimate the amount of variance in reading [mathematics] achievement that was within (Level 1) and between (Level 2) schools, the HLM analysis began with a one-way random-effects ANOVA model. This initial unconditional model—the null model—estimated an average reading [mathematics] scale score across schools of 43.54 [38.36]. The between-school variation, $\tau_{00} = 46.30$ [85.86] was statistically significant, $\chi^2(21) = 2{,}663.55$, $p < .001$ [$\chi^2(21) = 3{,}452.16$, $p < .001$], indicating significant differences among schools in their mean achievement levels. The estimated intraclass correlation coefficient of .13 [.17] indicated that about 13% [17%] of the variance in reading [mathematics] scores was due to schools. The within-school variation was $\sigma^2 = 303.29$ [423.35]. A high reliability statistic of .99 [.99] indicated that the model's estimation of the means provided reliable indicators of the true school means.

Next, we estimated the school-level model (means-as-outcomes regression model) with no student-level predictors. This model regressed average reading [mathematics] achievement on school poverty rate. The results of the model indicated a significant association between school poverty and mean reading [mathematics] achievement, $t(20) = –8.02, p < .001$ [$t(20) = –9.94, p < .001$]. School poverty rate accounted for 75% [82%] of the between-school variability in schools' mean reading [mathematics] achievement.

The next step was to estimate a student-level model (the random-coefficient regression model). This model regressed reading [mathematics] achievement on gender, age, current ELL, former ELL, and SES status. No school predictors entered this model. The results indicated that all five [four for mathematics; gender was not significant, $p = .623$] student-level variables were significantly related to reading [mathematics] achievement ($p < .001$). Because gender was not significant for mathematics, it was removed from subsequent mathematics models, but retained for reading models. Introducing gender (reading only), age, current ELL, former ELL, and SES status into the model explained 15% [15%] of the within-school variability.

Finally, a two-level model of academic achievement in reading [mathematics] was estimated. The Level 1 model included the intercept and five [four] slopes: gender (reading only), age, current ELL, former ELL, and SES. School poverty rate served as a predictor of Level 1 means and current ELL, former ELL, and SES slopes in the Level 2 model. Only variables that predicted reading [mathematics] scores in the full model at $p < .05$ were retained for the final, explanatory model. Table 2 displays estimates produced by the final models.

Table 2 Final models of reading and mathematics achievement

	Final reading model		Final math model	
Fixed effects	Coefficient (*SE*)	*t* Ratio	Coefficient (*SE*)	*t* Ratio
School mean (β_{0j})				
Base (γ_{00})	51.14 (1.05)	48.62***	44.33 (1.26)	35.28***
POVERTYR(γ_{01})	−0.28 (0.05)	−5.49***	−0.44 (0.06)	−6.85***
Gender (β_{1j})				
Base (γ_{10})	−5.58 (0.24)	−23.52***	ns	
Age (β_{2j})				
Base (γ_{20})	−0.97 (0.12)	−8.29***	−2.84 (0.14)	−20.50***
Current ELL slope (β_{3j})				
Base (γ_{30})	−10.30 (1.76)	−5.84***	−9.98 (1.66)	−6.01***
POVERTYR (γ_{31})	ns		0.20 (0.08)	2.46*
Former ELL slope (β_{4j})				
Base (γ_{40})	9.65 (0.76)	12.64***	9.52 (0.96)	9.87***
POVERTYR (γ_{41})	0.14 (0.04)	3.64**	0.12 (0.05)	2.37*
SES slope (β_{5j})				
Base (γ_{50})	−8.65 (1.06)	−8.15***	−11.36 (1.24)	−9.17***
POVERTYR (γ_{51})	0.12 (0.05)	2.19*	0.19 (0.06)	2.95**
Random effects	Variance (*df*)	Chi-square	Variance (*df*)	Chi-square
Between-school means (τ_{00})	22.58 (20)	772.77***	32.76 (20)	813.11***
Current ELL slope (τ_{30})	40.74 (20)	115.72***	27.36 (20)	50.17***
Former ELL slope (τ_{40})	0.78 (20)	20.50	3.65 (20)	12.26
SES slope (τ_{50})	22.76 (20)	423.77***	30.85 (20)	417.69***
Within-schools (σ^2)	257.27		360.05	

Note. Values are reported in scale score points. ns = not significant.
*$p < .05$.
**$p < .01$.
***$p < .001$.

Final Explanatory Models

Estimating the Means

The final explanatory models were specified in order to—after controlling for gender (reading only) and age—separately estimate the unique and interaction effects of students' English proficiency (NES, which served as the baseline from which to derive the reference group, current ELL, and former ELL), SES, and school poverty rates on individual reading [mathematics] achievement. Table 2 displays estimates produced by the final models. The school mean base value (γ_{00}) indicated the average reading [mathematics] score for the reference student. This reference student was a high SES, native English-speaking, average age female student [combined gender for mathematics] attending a school with an average poverty rate (see Table 1 for the coding scheme). This student's mean reading [mathematics] score was 51.14 [44.33] (γ_{00}).

School poverty rate and the average reading [mathematics] score were negatively related (γ_{01}), with each percentage of higher poverty rate predicting a decrease in reading [mathematics] scores of 0.28 [0.44] scale points. For example, in the highest poverty school in the sample (a school that had 37% more students eligible for free or reduced-price lunch than the average school), the average reading [mathematics] score was lower by 10.4 [16.3] points compared to average-poverty schools. The average reading [mathematics] score was higher by about 9.3 [14.5] points in the lowest poverty school in the sample (a school that had 33% fewer students eligible for free or reduced-price lunch than the average school) compared to average-poverty schools.

Estimating the Slopes

Results for the final models in Table 2 also display the unique effects associated with student-level predictors and their interactions with school poverty. On average, reading performance of boys was lower than that of girls by 5.58 points (γ_{10}), whereas in mathematics there was no significant difference by gender. One year increase in age was associated with a 0.97 [2.84] point decline in students' reading [mathematics] scores (γ_{20}). After controlling for all other variables in the models, the reading [mathematics] score of a current ELL student was lower than that of a NES student by 10.30 [9.98] points (γ_{30}). There was no significant interaction between current ELL status and school poverty rate for reading; the interaction, however, was positive and significant for mathematics, predicting a 0.20-point (γ_{31}) decrease in the current ELL mathematics disadvantage per each 1% increase in school poverty rate. In other words, the current ELL gap in mathematics (but not in reading) was smaller in higher poverty schools as compared to that in lower poverty schools.

Notably, on average, former ELLs outperformed their NES counterparts in reading [mathematics] by 9.65 [9.52] points (γ_{40}) and outperformed their current ELL peers by 19.95 [19.50] points ($|\gamma_{30}| + |\gamma_{40}|$). School poverty had a small positive interaction effect of 0.14 [0.12] points on former ELLs' reading [mathematics] scores (γ_{41}); that is, each 1% increase in school poverty rate predicted a slight score increase by 0.14 [0.12] points for former ELLs.

Table 3 Predicted average reading and mathematics scale scores by gender, language group (NES, current ELL, and former ELL students), and individual SES

	Reading scores		Mathematics scores[a]	
Student characteristics	High SES	Low SES	High SES	Low SES
Female				
Native English-speakers (NES)	51.14	42.49	44.33	32.99
Current ELL	40.84	32.19	34.35	23.01
Former ELL	60.79	52.14	53.85	42.51
Male				
Native English-speakers (NES)	45.56	36.91		
Current ELL	35.26	26.61		
Former ELL	55.21	46.56		

Note. These estimates are based on an average-age student (about seventh grade) enrolled in a school that is average on poverty rate.
[a]Because the gender effect was significant only for reading, the predicted scores in mathematics are reported for both genders combined.

These results suggest that former ELLs did not lose their advantage in higher poverty schools. In fact, relative to the school means, former ELL advantage was slightly higher in higher poverty schools than it was in lower poverty schools.

Low-SES students had an 8.65 [11.36]-point (γ_{50}) disadvantage in reading [mathematics] scores in comparison to their more economically advantaged peers. This effect was slightly mitigated by school poverty rate, predicting a slight increase by 0.12 [0.19] points (γ_{51}) in reading [mathematics] scores for each 1% increase in school poverty rate. Table 3 summarizes middle graders' predicted reading [mathematics] scores by gender (reading only), language status, and individual SES; these scores were computed based on the estimates produced by the final models. Notably, the predicted scores of low-SES former ELLs are comparable to those of high-SES NES students.

Practical Significance

The practical significance of these results was evaluated in three ways. First, the standardized effect sizes for group means differences (Cohen's *d*) were computed by using the following formula: $d = \gamma_{p0}/(\tau_{00} + \sigma^2)^{1/2}$ (Spybrook, Raudenbush, Congdon, & Martínez, 2009, p. 57), where τ_{00} and σ^2 are Level 1 and Level 2 variance components of the null models and γ_{p0} are the pairwise group mean differences from the final models computed such that group differences arising from other relevant characteristics (i.e., age, gender, SES, and school poverty) are partialed out. Because the variance components computed by HLM represent the total variance in the dependent variables in the sample, this necessarily would be larger than if variance of only two comparison groups were pooled at a time. Thus, the effect sizes computed below underestimated the true effect sizes and should be interpreted as minimum effect sizes.

The effect sizes and 95% confidence intervals (CIs) for reading favored former ELLs in comparison to both NES and current ELL students, $d = 0.52$, 95% CI [0.43, 0.61] and $d = 1.07$, 95% CI [0.94, 1.20], respectively. The result were similar for mathematics, favoring former ELLs in comparison to NES students, $d = 0.42$, 95% CI [0.33, 0.51], and in comparison to current ELLs, $d = 0.86$, 95% CI [0.73, 0.99]. The effect sizes between NES and current ELL students favored NES students and were $d = 0.55$, 95% CI [0.47, 0.63] and $d = 0.44$, 95% CI [0.36, 0.52] for reading and mathematics, respectively. According to the conventional cutoff criteria proposed by Cohen (1977, 1988, as cited in Lipsey & Wilson, 2001), these effect sizes were medium (cutoff range: >0.20) to large (cutoff range: ≥ 0.80) in magnitude.

A second way of evaluating practical significance followed Raudenbush and Bryck's (2002) recommendation to estimate the increase in proportion of variance explained at each level from the unconditional to the final model.[10] At Level 2, between-school variation in (a) average reading [mathematics] achievement, (b) former ELL effect, and (c) SES effect was substantially reduced by 61%, 95%, and 21% [70%, 64%, and 32%], respectively. In fact, the remaining residual (unexplained) variance in the former ELL effect was not significantly different from zero for reading [mathematics], $\chi^2(20) = 20.50$, $p = .427$ [$\chi^2(20) = 12.27$, $p > .500$]. Notably, for current ELLs, introducing school poverty rate as a Level 2 predictor reduced the amount of unexplained variance in mathematics scores by 32% but increased the unexplained variance in reading scores by 16%. The latter effect may be attributed to a nonsignificant interaction between current ELL status and school poverty rate. A sizable amount of between-school variance for average reading and mathematics achievement, SES, and current ELL effects remained unexplained.

At Level 1, the final model explained 15% [15%] of within-school variance in reading [mathematics] achievement, suggesting that it may be fruitful to explore other student-level variables. In addition, classroom variables (e.g., teacher ESL preparation and experience, teacher instructional approaches, instructional materials available, etc.) not included in this study may account for differences across student groups. Future research investigating a three-level HLM (including the classroom as the middle level) may be able to parse out some of this variance.

A third way to evaluate practical significance is to review the magnitudes of the scale score differences among groups in comparison to the assignment of one of the four proficiency levels by the state testing program. Because percentages of students falling within each proficiency level is a key component of the annual judgment of school performance, this lens applied to practical significance would likely to be of particular interest to school districts. For these results, differences in order from highest to lowest in scale scores between former ELLs, NES, and current ELLs tended to be approximately 10 scale scores points between each group for both reading and mathematics (see Table 2). This magnitude of differences can be considered in comparison to a proficiency level scale score span of approximately 20 points each. Thus, these results suggest that the three groups were each separated by approximately half of a proficiency level on average—a practical effect size that might reasonably be claimed as large. Aggregating across comparison groups, 78% [66%] of former ELLs scored Proficient or Distinguished on KCCT reading [mathematics] compared to 62% [51%] and 23% [15%] of NES and current ELLs, respectively (see Appendix C).

Discussion

Results of HLM analyses indicated that after controlling for gender, age, and student- and school-level SES, former ELLs significantly outperformed NES students and current ELLs on the state reading and mathematics tests. Additionally, former ELLs did not lose their advantage in higher poverty schools. In fact, those in higher poverty schools did slightly better than did their former ELL peers in lower poverty schools, relative to their respective school means. This may be due to higher levels of linguistic diversity in high poverty schools or to a greater amount of additional educational resources available in high poverty schools through Title I funding.

Former ELLs' outperforming current ELLs is an expected result consistent with Cummins's (1979, 2000) lower level TH holding that students attaining this threshold would avoid academic disadvantages associated with limited access to educational environment through a not-yet-fully-developed L2. Attaining the lower level threshold was operationalized in this study as reclassification from current ELL ("partial" bilinguals) to former ELL (at least "dominant" bilinguals) based on standardized English proficiency measures. This result is encouraging, implying that efforts to bring current ELL students to English proficiency may result in a substantial jump in standardized test scores once the language of schooling and testing has been adequately mastered.

The significantly higher mathematics and reading achievement of the former ELL group compared to that of NES students deserves further exploration. There are two primary hypotheses that we explore below. One hypothesis is that the reclassification process from current ELL to former ELL actually functions as an academic ability self-selection process (see below). A second, alternative hypothesis is that reclassification would identify ELL students from the full range of academic abilities whose academic success may be attributed, at least in part, to capitalizing on bilingual cognitive processing advantages upon reaching at least "dominant" bilingual status. This second hypothesis rests on a refinement of Cummins's (1979, 2000) TH. Below, we argue why the self-selection hypothesis is undermined in this study and argue how this study offers support for a refinement of the TH.

Self-selection Hypothesis Counterargument

One plausible explanation for the strong academic achievement of the former ELL group is that former ELLs are a self-selected group whose greater academic ability underlies their academic success (both in terms of reaching English proficiency reclassification and in terms of academic achievement). This self-selection hypothesis would suggest that the reclassification status splits the total ELL population in two groups including lower (current ELLs) versus higher (former ELLs) ranges of academic ability. Under this assumption, the comparison of NES students and former ELLs would be confounded because the NES group would comprise a total range of academic abilities. Indeed, when we fitted HLM models to compare academic achievement of the combined, current- and former-ELL (total ELL) student sample with that of NES students, the results favored the total ELL group but were not statistically significant either for reading ($\gamma_{\text{total ELL}} = 0.44$, $SE = 1.52$, $p = .776$) or for mathematics ($\gamma_{\text{total ELL}} = 1.34$, $SE = 1.31$, $p = .316$), after controlling for age, gender, SES, and school poverty (hence, there is no achievement gap when considering the full ELL population to include those who exited ELL status). On the surface, these results would support that the total ELL group is equivalent to the NES group in academic ability, in spite of the fact that the ELL group had to demonstrate its academic ability in an L2. These results, however, may only tell us that the *academic achievement* (i.e., test scores)—not the underlying *content knowledge* or *academic ability* (i.e., ability that may not be fully expressed in an L2)—had a comparable distribution across the two (total ELL and NES) groups. We offer three arguments against the self-selection hypothesis: one empirical argument from our study's results and two arguments based on how the study constructs were measured.

Counterargument #1: Years in L2 Environment

The first (empirical) counterargument is based on the number of years the two ELL groups (former and current) had spent in the L2 schools. If the self-selection hypothesis were strictly true, then one might reasonably assume that the two groups created by the selection process would differ only in academic ability, not any other variable. However, our data showed that former ELLs had a significant advantage over current ELLs with regard to the amount of schooling in and through English. The former ELL group's average time in the U.S. schools was significantly higher than that of the current ELL group (6.63 and 4.72 years, respectively; see Appendix B). This is consistent with research documenting ELLs' requiring up to 7 (or more) years to reach national benchmarks on standardized tests in content areas (Collier & Thomas, 1989; Thomas & Collier, 2002); comparable time-to-academic-proficiency-benchmarks results (4–7 years) were reported for NES students schooled through two languages (Collier, 1989, as cited in Bigelow, 2010). This suggests that the identification process for former ELL status did validly distinguish students based on the amount of instruction in and through the L2—which research (Ardasheva, 2010; Carhill et al., 2008; Páez, 2002) has identified as a strong positive predictor of L2 proficiency among ELLs—and not necessarily based on academic ability, particularly because other differentiating characteristics (i.e., age, gender, and SES) were statistically controlled. This conclusion is reinforced by ACCESS scores of current and former ELLs (see Appendix A). The current ELLs' median performance (5.2 on a 6-point scale) on the speaking component of ACCESS was not substantially lower than that of former ELLs (6.0), whereas the current ELLs' average performance was substantially lower on the reading (2.9 vs. 5.9) and—to lesser extent—writing (3.0 vs. 4.0) subscales. This implies that, on average, the current ELL group had adequate general cognitive abilities to make substantial progress toward English proficiency in speaking, and—because the ability to comprehend, interpret, and produce school-based texts is prerequisite for academic success across all school subjects (Cummins, 1981, 2008; Schleppegrell, 2004)—stronger future proficiency in English literacy skills may result in substantial jumps in academic achievement scores for this group. The 2-year gap in time in the U.S. schools (rather than underlying academic ability) could plausibly account for the lagging proficiency in English literacy skills. This is consistent with research documenting ELLs' requiring 2–5 years to develop mastery of English oral (Hakuta et al., 2000) as well as syntactic, morphological, and pragmatic skills (MacSwan & Pray, 2005) and up to 7 (or more) years to reach literacy skills comparable to NES student averages on standardized tests in reading (Collier &

Thomas, 1989; Hakuta et al., 2000; Thomas & Collier, 2002; see also a review in Cummins, 2000) suggesting that the identification process for former ELL status did validly distinguish students based on L2 proficiency—which is contingent on the amount of L2 instruction—and not necessarily based on academic ability.

Counterargument #2: Selection Measure Based on Language and Not Academic Content

Consistent with counterargument #1 that the selection process resulted in ELL groups characterized by English language proficiency only, the nature of the measures used to categorize former ELLs are language based and not academic content based. Whereas some school districts use multiple criteria for exiting students from current ELL status (e.g., academic achievement in core content areas plus English proficiency data; Solórzano, 2008), the district in this study made reclassification decisions based on measures of English proficiency only, either LAS or ACCESS, both measuring *language-specific knowledge* (e.g., English vocabulary, phonology, syntax). Although the content of the ACCESS items is provided by academic subject matter, student responses are evaluated against linguistic—rather than academic knowledge—performance criteria (i.e., vocabulary usage, structure complexity, and language control; WIDA, 2008). Poor quality of LAS in predicting ELLs' academic achievement has been recognized both by educators (Faltis & Coulter, 2007) and by researchers (see Solórzano, 2008) and the majority of former ELLs in the sample ($n = 336$) were reclassified based on LAS. Although both LAS and ACCESS included measures of English reading skills—suggesting a possible overlap with KCCT reading scores[11]—neither test measured students' mathematics knowledge. Thus, without some additional academic knowledge testing either in the students' strongest language (Solórzano, 2008) or through alternative assessment (Abedi, 2004; Kopriva et al., 2009; Koretz, 2008), no valid claims regarding student general academic knowledge can be drawn from the ELL reclassification tests. This suggests that reclassification based on language proficiency assessment can allow us to validly differentiate between low- and high-English-proficiency students, not between low- and high-academic-ability or -academic-knowledge students.

Counterargument #3: Limitations of English-Language Academic Achievement Tests to Portray Academic Ability of ELLs

The standardized academic achievement tests (KCCT in this study) are limited in terms of providing a complete and accurate picture of the actual state of current ELL academic knowledge and ability (Abedi, 2004, 2007; Koretz, 2008;

Solórzano, 2008), calling into question a basic feature of the self-selection hypothesis that posits that the current ELL group is significantly lower in academic ability compared to the former ELL group. As argued in the introduction, current ELLs lack the fine-tuned receptive and productive English language skills that would allow them to discriminate between multiple-choice items and to provide constructed responses demonstrating their content knowledge (Kopriva, 2009; Solórzano, 2008). Their lower performance on standardized tests may reflect the lack of proficiency in the language of testing, or the lack of content knowledge and skills, or a combination of the two (Goldenberg, 2008). A substantial body of empirical evidence—discussed throughout the article—has indicated that language proficiency interferes with academic achievement scores on tests administered in the L2. Reading comprehension research, for example, has shown that even in the presence of a substantial relationship between L1 and L2 academic skills, low levels of L2 proficiency "short circuit" or limit (Clarke, 1979, 1980; Lee & Schallert, 1997; Schoonen et al., 1998) the extent to which language learners can draw on their academic ability to support their L2 performance. Further, there is evidence (Lee & Schallert, 1997; Mahon, 2006) to suggest that a substantially greater proportion of variance in L2 performance may be accounted for by L2 proficiency than by underlying academic ability measured through students' L1s.[12] Moreover, evidence from academic achievement research indicates that minimizing language demands or providing alternative assessment may improve current ELL students' test scores (e.g., Abedi & Lord, 2001; Kopriva et al., 2009). To reiterate, these findings suggest that (a) learners need to develop high levels of L2 proficiency before they can capitalize on their (L1) academic skills (Lee & Schallert, 1997; Schoonen et al., 1998); (b) L2 proficiency may play a substantially larger role in L2 performance than learners' underlying academic skills (Lee & Schallert, 1997; Mahon, 2006); and (c) low levels of L2 proficiency prevent ELLs from demonstrating their true levels of academic ability and knowledge (Kopriva et al., 2009), developed either through L1 or L2 schooling.

In sum, based on the three counterarguments, we argue that although ELL reclassification distinguished between low- and high-English-proficiency students (particularly in the literacy domains), the selection process to categorize former ELLs did not necessarily discriminate the two groups based on underlying academic ability. Thus, in contrast to the self-selection hypothesis, English proficiency—rather than academic ability—is more likely to be the key variable that distinguished former and current ELLs. However, confidently accepting or rejecting the self-selection hypothesis would require additional research, perhaps involving nonverbal measures of intelligence for both NES and ELL groups or measures of academic achievement through alternative assessment such as described in Kopriva et al. (2009).

Alternative Argument: Refinement of Bilingual Threshold Hypothesis

As an alternative to the self-selection hypothesis for explaining the former ELL academic achievement score advantage over NES, we offer a refinement of Cummins's (1979, 2000) TH. The former ELL advantage over NES suggests that some factors in addition to adequate proficiency in the language of schooling—a characteristic of both NES and former ELLs—may have contributed to the academic performance of former ELLs. An empirical body of work (e.g., Bialystok, 1986, 1999; Kaushanskaya & Marian, 2009; Kormi-Nouri et al., 2008; see also Adesope et al., 2010) consistent with this finding is that higher achievement may be attributed, in part, to *cognitive processing benefits* associated with bilingualism.

As noted in the literature review, two potential sources of bilingual advantages have been advanced in the literature—namely, biliteracy and cognitive processing advantages. Whereas evidence (Clarke, 1979, 1980; Lee & Schallert, 1997; Proctor et al., 2010) has indicated that the biliteracy advantage (transfer) would require substantial levels of literacy skills in both languages, research on preschool bilinguals (e.g., Bialystok, 1986, 1999; see also Adesope et al., 2010) has suggested that the cognitive processing advantage may be independent of (or, at least, dependent to a lesser degree on) biliteracy skills. In other words, evidence from the latter set of studies has suggested that cognitive processing advantages associated with bilingual experiences—at least when it comes to symbolic representation, attentional control, and problem solving, all of which are likely to have an impact on students' academic functioning—may arise from children's developing oral proficiency in two languages, often through a mere exposure to two languages on an everyday basis (e.g., L1 at home and L2 elsewhere; Bialystok, 1986, 1999). This is consistent with the hypothesis (e.g., Adesope et al., 2010; Bialystok, 1999; Costa et al., 2008; see also the review of supporting empirical evidence from neurology research in Mindt et al., 2008) holding that bilingual cognitive processing advantages may be due to human brain's adopting to the ongoing need for operating in the dual-language environments that results in the "added practice" in exercising cognitive control in selecting or inhibiting one of the two active languages and may ultimately lead to greater efficiency in exercising cognitive control among bilinguals.

The present evidence suggests that some refinements in the TH, particularly as it applies to ELLs, are needed; that is, our study findings, when taken together with past theoretical and empirical evidence, suggest that *oral proficiency* *in two languages*—not necessarily *biliteracy skills* as originally proposed by Cummins (1979, 2000)—developed either through language exposure or through language instruction may be a sufficient condition for cognitive processing benefits of bilingual experiences to become available to the student. Consistently with Cummins, however, such cognitive processing benefits would translate into higher academic performance only once the language of schooling, particularly its literacy domains, has been adequately developed. This suggests that the academic benefits associated with bilingualism may be available at *both* thresholds—in contrast to only at the higher threshold, as originally proposed by Cummins—by virtue of students gaining access to and being able to capitalize on different aspects of bilingual advantages (i.e., cognitive processing benefits at the lower threshold and biliteracy advantage at the higher threshold). Additionally, the evidence we offer opens up the possibility that the academic benefits associated with bilingualism may be *additive* in nature, with cognitive processing benefits strengthening student academic achievement at both the lower and higher level thresholds and biliteracy benefits further strengthening student academic achievement at the higher level threshold. The latter hypothesis, however, is speculative and would require additional research.

The academic achievement advantage of former ELLs over NES students may hypothetically be attributed, in part, to a presence of "balanced" bilinguals among former ELLs, a framework consistent with the higher level TH (Cummins, 1979, 2000). The presence of "balanced" bilinguals, however, could not be formally verified in this study due to a lack of standardized L1 literacy skills' measure and—unless ELL students come with a strong literacy foundation in their L1s or receive continued support in L1 literacy skills at home—is less likely in nonbilingual educational environments (Proctor et al., 2010), such as the one in this study, particularly because the majority of former ELLs (76%) received most of their formal education (about 6 or more years) in the U.S. schools.

In sum, an explanation consistent with the results of our study is that cognitive processing benefits may be available to both current and former ELLs, but may translate into academic benefits as measured by achievement tests upon reaching the lower level threshold, once the language of schooling and testing has been mastered. This may be particularly relevant for ELLs schooled in nonbilingual educational contexts because of the noted asymmetry in language skills development under this instructional model, which (a) is often associated with L1 literacy skills below grade-level norms (Proctor et al., 2010) and (b) may potentially limit opportunities for transfer of academic and literacy skills across languages (Clarke, 1979, 1980; Lee & Schallert, 1997; Proctor et al., 2010; Schoonen et al., 1998).

Conclusion

Historically, bilingual experiences of language minority students have been linked to cognitive, academic, and social disadvantages for the child (see discussions by Cummins, 1979; Lambert, 1978; Peal & Lambert, 1962). The two linguistic systems in one person were perceived as competing for cognitive resources—a theoretical position known as the cross-language interference hypothesis—thus leading to "mental confusion" and poor academic performance. However, as noted by Cummins (1979) and Lambert (1978), early research on bilingual children often failed to control for relevant background characteristics (e.g., SES, levels of L2 proficiency). By building on Cummins's (1979, 2000) bilingual TH, this study contributed to a more recent body of research on the correlates of bilingualism, advanced our understandings of the mediating role that L2 proficiency levels play in ELLs' academic development, addressed a deficit view of language minority students, and revealed information that has potential significance for policy makers and for future research.

Over time, the research community has both criticized and supported Cummins's (1979, 2000) TH. This study addressed some, but not all, of the methodological issues raised in the literature (MacSwan, 2000; see also Takakuwa, 2005) and thus offers several methodological as well as a theoretical advancements for studying this theory. First, this study used multilevel statistical procedures that allowed controlling for SES effects both at the student and school levels. Second, the study minimized the concerns raised about improper reliance on statistical significance by reporting effect sizes and confidence intervals. Third, this study conducted a more rigorous test of Cummins's lower level TH by comparing academic achievement of ELLs above and below this L2 proficiency threshold operationally defined in an *absolute* sense—as opposed to in a *relative*, *within-group* sense—as the reclassification from current to former ELL status based on standardized English proficiency measures. The absolute definition of L2 proficiency used in this study enhances generalizability of the findings and would allow for replication of research, at least within the United States.[13]

Finally, by arguing in favor of distinguishing between oral and literacy skills, this study was able to propose some refinements to the TH as well as a more nuanced perspective on the relationships between language proficiency and cognitive and academic development among bilinguals; that is, we propose that (a) oral proficiency in two languages may be a sufficient condition for cognitive processing benefits of bilingual experiences to become available to the student and (b) the academic benefits associated with bilingualism may be available at both thresholds by virtue of students' gaining access to and being able to capitalize on different aspects of bilingual advantages (i.e., cognitive processing benefits at the lower threshold and biliteracy advantage at the higher threshold).

A concern that this study was not able to address is the question of causality between bilingualism and academic achievement; that is, because we do not have longitudinal L1 and L2 proficiency data, this study is unable to make a causal claim. Instead, these results support that a correlation exists between at least "dominant" bilingualism and an academic achievement advantage, suggesting that—in contrast to the TH as originally formulated by Cummins (1979, 2000)—academic benefits associated with bilingualism may be available at the lower threshold of bilingual proficiency. The existence of this correlation—as well as recent theoretical and empirical advancements in bilingual research (e.g., Adesope et al., 2010; Mindt et al., 2008)—suggests possible refinements for the TH (see above) and argues that research with access to appropriate data may be able to more fully test the hypothesis.

Further, similar to Kim and Herman's (2009) three-state study and NYCDE (2009) report, this study highlights the strong academic achievement attained by former ELLs in comparison to NES students. These convergent findings from various contexts suggest the possibility of stable, replicable patterns in former ELL students' high achievement, and suggest that publicized achievement gaps between (current) ELLs and NES students may be misrepresenting how ELLs eventually fare. If state and national policies to evaluate academic success of ELL students and ELL programs were to publicly include former ELL scores along with current ELL scores, this might offer a more representative picture of ELL student outcomes. Additionally, these results call for more fair assessment practices for current ELLs (see Kopriva et al., 2009), provide consequential validity support for using reclassification into former ELL as an adequate criterion for exiting ELL students from language support services, and suggest that academic and language support programs that lead to English proficiency may be effective for enabling ELLs to attain higher academic achievement. Yet, about 25% of ELLs across the nation fail to reach the English proficiency benchmark (Quality Counts, 2009).

One of the proposed approaches to supporting ELLs' language, and ultimately, academic outcomes has been for all educators, including content-area educators, to become language teachers for ELLs (Lucas, Villegas, & Freedson-Gonzalez, 2008; Scarcella, 2003). Theory (DeKeyser, 2003; Ellis, 1994; Hulstijn, 2005; Robinson, 1995; Schmidt, 1990) and research (see reviews by DeKeyser, 2003; Norris & Ortega, 2000; Spada & Tomita, 2010) suggest that L2 development benefits from varied forms of explicit language instruction, including in content-area classrooms (Zwiers, 2006, 2007). Yet, according to the 2009 Quality Counts report, currently only 11 states offer teacher incentives for earning an ESL license and/or endorsement and only three states require that all preservice teachers demonstrate their competence in working with ELLs in order to earn a teaching license. Another approach is to offer—when feasible—bilingual education programs, which not only produce results superior to English-only programs (e.g., Rolstad et al., 2005; Slavin & Cheung, 2005), but also have an added advantage of developing and maintaining L1 literacy skills among ELLs (Proctor et al., 2010).

Additional research is needed to further explore academic achievement of former ELLs—when feasible, disaggregated by types of language support programs received in the past—in (a) higher grades in which the linguistic demands of subject areas are, arguably, more challenging than in middle grades (Genesee et al., 2005) and (b) other content areas in addition to reading and mathematics. Furthermore, longitudinal research that tracks individual ELL student scores over time, including both before and after they achieve adequate English proficiency on the state-designated assessments, and controls for nonverbal intelligence and L1 literacy and oracy will permit for a more rigorous test of the TH in the future.

Revised version accepted 1 February 2011

Notes

1 Here, Cummins (2000) recognized the role of *bilingual* versus *monolingual* instructional contexts. “In the former case,” he wrote, “the hypothesis would argue that a certain degree of academic proficiency in both languages is required whereas in the latter, attaining sufficient proficiency in just the L2 (the language of school instruction) might suffice” (p. 175).

2 Cummins (1981, 2000, 2008) defined *academic language proficiency* as literacy-related skills associated with school-based tasks and contrasted this concept with that of *social language proficiency*, the latter including basic pronunciation, vocabulary, and grammar skills sufficient to maintain social interactions.

3 Although the question of whether oral language proficiency represents a coherent and meaningful construct that is distinguishable from literacy skills remains a matter of debate (partly because of high correlations between measures of oral

vocabulary and reading comprehension; see a discussion in Cummins, 2009), many researchers (e.g., August & Shanahan, 2006; Bialystok, 1986, 1999) found the distinction useful, particularly with regard to studying young, preliteracy or emergent-literacy bilinguals.

4 Cummins (2000) referred to this phenomenon as a *linguistic threshold* "below which L1 reading will exert less impact on L2 reading" than L2 proficiency will (p. 196) and critiqued Lee and Schallert's (1997) study—as well as other studies within this line—on the grounds that the constructs of L2 proficiency and L2 reading (both administered through a "written" mode) are not independent, essentially assessing the same construct of academic literacy skills (p. 197). Yet, the operationalization of L2 proficiency and L2 reading in Lee and Schallert did allow one to define skills underlying the two constructs as nested (Cummins's point), yet highly distinguishable in terms of complexity, with the former measure capturing the knowledge of discrete linguistic elements (vocabulary, grammar) and the latter measure capturing the ability to apply one's knowledge of discrete linguistic elements to derive meaning from text (i.e., the meaning-generating ability dependent on sophisticated knowledge of semantics and written discourse structures as well as the ability to engage general knowledge and higher order thinking skills).

5 This decision was made in order to exclude students whose high ELP scores upon school enrollment suggested that these students were either born in or had been brought to the United States at a very young age and for whom English may have become their L1 (MacSwan, 2000); the average within-school IFEP sample size was seven (range: 0–20).

6 Schools with at least one current or former ELL within their bodies were not excluded because HLM has the advantage of "borrowing strength" from the entire sample in estimating individual school equations (Raudenbush & Bryk, 2002, pp. 6–7).

7 In the 2011–2012 school year, the district will adopt a new set of standardized tests aligned to national content-area standards.

8 Centering is a procedure that allows to set a meaningful baseline (comparison) point to a scale that does not have a natural or meaningful value of zero; group mean centering sets the zero at the within-school (group) mean and grand mean centering sets the zero at the between-school (grand) mean (Raudenbush & Bryck, 2002).

9 Here, the slope is a coefficient that estimates an average, between-school effect of each of the five student-level predictors (GENDER: $p = 1$; AGE: $p = 2$, and so forth) on the mean KCCT score in reading or mathematics.

10 The following formulas served to estimate the proportion of variance explained: (a) (σ^2 unconditional model $-\sigma^2$ final model)/σ^2 unconditional model (Level 1) and (b) (τ_{qq} random-coefficient regression model $-\tau_{qq}$ final model)/τ_{qq} random-coefficient regression model (Level 2).

11 The KCCT reading, however, assessed skills sufficiently distinct—except for what KCCT defines as "foundational" skills (e.g., suffixes, prefixes, synonyms)—from those assessed by LAS and ACCESS—namely, students' *content-specific* (Reading/Language Arts) *knowledge* (i.e., text interpretation, evaluation, and response skills; see Method section).

12 Footnote 4 summarizes Cummins's (2000) critiques of this line of research.

13 Although reclassification criteria vary from state to state (Solrzano, 2008), there is a push for the standardization of the procedure on the federal level (Notice of Final Interpretations, 2008) and, currently, 22 states and the District of Columbia have adopted ACCESS as their ELP measure.

References

Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher, 33,* 4–14.

Abedi, J. (2007). English language proficiency assessment and accountability under NCLB Title III: An overview. In J. Abedi (Ed.), *English language proficiency assessment in the nation: Current status and future practice* (pp. 3–13). Davis: University of California, Davis School of Education.

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14,* 219–234.

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualisms. *Review of Educational Research, 80,* 207–245.

Andreou, G., & Karapetsas, A. (2004). Verbal abilities in low and highly proficient bilinguals. *Journal of Psycholinguistic Research, 33,* 357–364.

Ardasheva, Y. (2010). *English language learners in focus: Predictors of English proficiency and academic achievement*. Unpublished doctoral dissertation, University of Louisville, Louisville, KY.

August, D., & Shanahan, T. (Eds.). (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ: Erlbaum.

Bialystok, E. (1986). Factors in the growth of linguistic awareness. *Child Development, 57,* 498–510.

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development, 70,* 636–644.

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science, 7,* 325–339.

Bialystok, E., McBride-Chang, C., & Luk, G. (2005). Bilingualism, language proficiency, and learning to read in two writing systems. *Journal of Educational Psychology, 97,* 580–590.

Bigelow, M. H. (2010). *Mogadishu on the Mississippi: Language, racialized identity, and education in a new land*. Malden, MA: Wiley-Blackwell.

Carhill, A., Suárez-Orozco, C., & Páez, M. (2008). Explaining English language proficiency among adolescent immigrant students. *American Educational Research Journal, 45,* 1155–1179.

Chomsky, N. (1959). A review of B. F. Skinner's "Verbal behavior." *Language,* 35, 26–58.

Clarke, M. A. (1979). Reading in Spanish and English: Evidence from adults ESL students. *Language Learning, 29,* 121–150.

Clarke, M.A. (1980). The short circuit hypothesis of ESL reading—or when language competence interferes with reading performance. *Modern Language Journal, 64,* 203–209.

Collier, V. P., & Thomas, W. P. (1989). How quickly can immigrants become proficient in school English? *Journal of Educational Issues of Language Minority Students, 5,* 26–38.

Cook, H. G., & Wilmes, C. (2007). *Alignment between the Kentucky Core Content for Assessment and the WIDA Consortium English language proficiency standards*. Madison: University of Wisconsin, Wisconsin Center for Education Research.

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition, 106,* 59–86.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49,* 222–251.

Cummins, J. (1981). Empirical and theoretical underpinnings of bilingual education. *Journal of Education, 163,* 16–30.

Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire*. Clevedon, UK: Multilingual Matters.

Cummins, J. (2008). BICS and CALP: Empirical and theoretical status of the distinction. In B. Street & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol. 2. Literacy* (2nd ed., pp. 71–83). New York: Springer.

Cummins, J. (2009). Literacy and English-language learners: A shifting landscape for students, teachers, researchers, and policy makers. [Review of the book *Developing reading and writing in second-language learners: Lessons from the Report of the National Literacy Panel on Language-Minority Children and Youth*]. *Educational Researcher, 38,* 382–384.

Dalton, S. (1979). Validation of the Language Assessment Scales. *Educational and Psychological Measurement, 39,* 1001–1003.

Datnow, A., Stringfield, S., & Castellano, M. (2005). School reform and the education of culturally and linguistically diverse students. In C. R. O'Donnell & L. A. Yamauchi (Eds.), *Culture and context in human behavior change: Theory, research, and application* (pp. 179–204). New York: Peter Lang.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 313–348). Oxford: Blackwell.
Diaz, R. M., & Klingler, C. (1992). Towards an exploratory model of the interaction between bilingualiasm and cognitive development. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 142–167). New York: Cambridge University Press.
Duncan, S. E., & De Avila, E. A (1988). *Language Assessment Scale (LAS-R/W)*. Monterey, CA: CTB/McGraw-Hill.
Duncan, S. E., & De Avila, E. A. (1994). *Language Assessment Scale (LAS-O)*. Monterey, CA: CTB/McGraw-Hill.
Ellis, N. C. (1994). Implicit and explicit language learning: An overview. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 1–32). San Diego, CA: Academic Press.
Faltis, C. J., & Coulter, C. A. (2007). *Teaching English learners and immigrant students in secondary schools*. Upper Saddle River, NJ: Merrill Prentice Hall.
Fromkin, V., Rodman, R., & Hyams, N. (2007). *An introduction to language* (8th ed.). Boston: Thomson Wadsworth.
Fry, R. (2007). *How far behind in math and reading are English language learners?* Washington, DC: Pew Hispanic Center.
Fry, R. (2008). *The role of schools in the English language learner achievement gap*. Washington, DC: Pew Hispanic Center.
Genesee, F., Lindholm-Leary, K., Saunders, W., & Christian, D. (2005). English language learners in U.S. schools: An overview of research findings. *Journal of Education for Students Placed at Risk, 10,* 363–386.
Goldenberg, C. (2008). Teaching English language learners: What the research does and does not say. *American Educator, 33*(2), 8–23.
Grubb, W. N. (2008). Multiple resources, multiple outcomes: Testing the "improved" school finance with NELS88. *American Educational Research Journal, 45,* 104–144.
Hakuta, K., Butler, Y., & Witt, D. (2000). *How long does it take English learners to attain proficiency?* (Policy Report 2000–1). San Francisco: University of California, Linguistic Minority Research Institute.
Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning. *Studies in Second Language Acquisition, 27,* 129–140.
Kao, G., & Thompson, J. S. (2003). Racial and ethnic stratification in education achievement and attainment. *Annual Review of Sociology, 29,* 417–442.
Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin and Review, 16,* 705–710.
KDE (Kentucky Department of Education). (2005). *Kentucky Core Content Tests: 2004 Technical Report*. Frankfort, KY: Author.

KDE (Kentucky Department of Education). (2008). *Monitoring former LEP students: Kentucky Department of Education Guidelines*. Frankfort, KY: Author.

KDE (Kentucky Department of Education). (2009). *No Child Left Behind Adequate Yearly Progress Report–2008*. Frankfort, KY: Author.

Kenyon, D. M. (2006). *Development and field test of ACCESS for ELLs®* (Tech. Rep. No. 1). Madison: University of Wisconsin, Wisconsin Center for Education Research.

Kim, J., & Herman, J. L. (2009). A three-state study of English learner progress. *Educational Assessment, 14,* 212–231.

Kim, Y.-H., & Jang, E. E. (2009). Differential functioning of reading subskills on the OSSLT for L1 and ELL Students: A multidimensionality model-based DBF/DIF approach. *Language Learning, 59,* 825–865.

Kopriva, R. (2009). Assessing the skills and abilities in math and science of ELLs with low English proficiency: A promising new method. *AccELLerate!! The quarterly newsletter of the National Clearinghouse for English Language Acquisition, 2*(1), 7–10. Retrieved November 11, 2009, from http://www.ncela.gwu.edu/files/accellerate/7/Accellerate_2_1.pdf

Kopriva, R. J., Bauman, J., Cameron, C., & Triscari, R. (2009). *Final research report Obtaining Necessary Parity Through Academic Rigor in science: A report to the U.S. Department of Education*. Retrieved November 11, 2009, from http://www.onpar.us/Reports/ONPARResearchReport.pdf

Koretz, D. (2008). *Measuring up*. Cambridge, MA: Harvard University Press.

Kormi-Nouri, R., Shojaei, R. S., Moniri, S., Gholami, A. R., Moradi, A. R., Akbari-Zardkhaneh, S., & Nilsson, L. G. (2008). The effect of childhood bilingualism on episodic and semantic memory tasks. *Scandinavian Journal of Psychology, 49,* 93–109.

Lambert, W. E. (1978). Cognitive and socio-cultural consequences of bilingualism. *Canadian Modern Language Review, 34,* 537–47.

Lasagabaster, D. (1998). The threshold hypothesis applied to three languages in contact at school. *International Journal of Bilingual Education and Bilingualism, 1,* 119–133.

Lazarin, M. (2006). *Improving assessment and accountability for English language learners in the No Child Left Behind Act*. (Issue Brief No. 6). Washington, DC: National Council of La Raza.

Lee, J-W., & Schallert, D. (1997). The relative contribution of L2 language proficiency and L1 reading ability to L2 reading performance: A test of the threshold hypothesis in an EFL context. *TESOL Quarterly, 31,* 713–739.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Lucas, T., Villegas, A. M., & Freedson-Gonzalez, M. (2008). Linguistically responsive teacher education: Preparing classroom teachers to teach English language learners. *Journal of Teacher Education, 59,* 361–373.

MacSwan, J. (2000). The Threshold Hypothesis, semilingualism, and other contributions to a deficit view of linguistic minorities. *Hispanic Journal of Behavioral Sciences, 22,* 3–45.

MacSwan, J., & Pray, L. (2005). Learning English bilingually: Age of onset of exposure and rate of acquisition among English language learners in a bilingual education program. *Bilingual Research Journal, 29,* 653–678.

Mahon, E. A. (2006). High-stakes testing and English language learners: Questions of validity. *Bilingual Research Journal, 30,* 479–497.

Mindt, M. R., Arentoft, A., Germano, K. K., D'Aquila, E., Scheiner, D., Pizzirusso, M., Sandoval, T. C., & Gollan, T. H. (2008). Neuropsychological, cognitive, and theoretical considerations for evaluation of bilingual individuals. *Neuropsychology Review, 18,* 255–268.

NCLB (No Child Left Behind Act of 2001). (2002). Pub. L. No. 107–110, 115 Stat. 1425.

NCPPHE (National Center for Public Policy and Higher Education). (2005). *Income of U.S. workforce projected to decline if education doesn't improve*. Retrieved September 20, 2009, from http://www. highereducation.org/reports/pa_decline/pa_decline.pdf#search=%22minorities%22

NYCDE (New York City Department of Education). (2009). *Diverse learners on the road to success: The performance of New York City's English language learners*. New York: Office of English Language Learners.

Norris, J. M., & Ortega, L. (2000). *Effectiveness of L2 Instruction: A research synthesis and quantitative meta-analysis. Language Learning, 50,* 417–528.

Notice of Final Interpretations, 73 Fed. Reg. § E8–24702 (2008).

Páez, M. (2002). Language and the immigrant child: Predicting English language proficiency for Chinese, Dominican, and Haitian students. *Dissertation Abstracts International, 62,* 3717A. (AAT No. 3034903)

Peal, E., & Lambert, W. E. (1962). The relation of bilingualism to intelligence. *Psychological Monographs: General and Applied, 76,* 1–23.

Proctor, C. P., August, D., Carlo, M., & Barr, C. (2010). Language maintenance versus language of instruction: Spanish reading development among Latino and Latina bilingual learners. *Journal of Social Issues, 66,* 79–94.

Quality Counts. (2009). Quality Counts 2009: Portrait of a population—How English-language learners are putting schools to the test [Special issue]. *Education Week, 28*(17).

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

Reardon, S. E., & Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grades. *American Educational Research Journal, 46,* 853–891.

Robinson, P. (1995). Attention, memory, and "noticing" hypothesis. *Language Learning, 45,* 283–331.

Rolstad, K., Mahoney, K., & Glass, G. V. (2005). The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational Policy, 19,* 572–594.

Scarcella, R. (2003). *Academic English: A conceptual framework* (Tech. Rep. No. 2003–1). Irvine: University of California, Linguistic Minority Research Institute.

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistic perspective*. Mahwah, NJ: Erlbaum.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11,* 129–158.

Schoonen, R., Hulstijn, J., & Bossers, B. (1998). Metacognitive and language-specific knowledge in native and foreign language reading comprehension: An empirical study among Dutch students in grades 6, 8 and 10. *Language Learning, 48,* 71–106.

Schrank, F. A., Fletcher, T.V., & Alvarado, C.G. (1996). Comparative validity of three English oral language proficiency tests. *Bilingual Research Journal, 20,* 55–68.

Slavin, R. E., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Educational Research, 75,* 247–284.

Solórzano, R. W. (2008). High stakes testing: Issues, implications, and remedies for English language learners. *Review of Educational Research, 78,* 260–329.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60,* 263–308.

Spybrook, J., Raudenbush, S. W., Congdon, R., & Martínez, A. (2009). *Optimal Design for longitudinal and multilevel research: Documentation for the "Optimal Design" software*. Retrieved November 10, 2009, from http://www.wtgrantfoundation.org/resources/overview/research_tools

Suárez-Orozco, C., Suárez-Orozco, M., & Todorova, I. (2008). *Learning a new land: Immigrant students in American society.* Cambridge, MA: Harvard University Press.

Takakuwa, M. (2005). Lessons from a paradoxical hypothesis: A methodological critique of the Threshold Hypothesis. In J. Cohen, K. T. McAlister, K. Rolstad, & J. MacSwan (Eds.), *ISB4: Proceedings of the 4th international symposium on bilingualism* (pp. 2222–2232). Somerville, MA: Cascadilla Press.

Thomas, W. P., & Collier, V. P. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement*. Santa-Cruz: University of California, Center for Research on Education, Diversity and Excellence.

Vollmer, G. (2000). Praise and stigma: Teachers' construction of the "typical ESL student." *Journal of Intercultural Studies, 21,* 53–66.

WIDA (World-Class Instructional Design and Assessment (WIDA™) Consortium). (2008). *Assessing comprehension and communication in English state-to-state for English language learners, ACCESS for ELLs®: Interpretative guide for score reports*. Madison: University of Wisconsin, Wisconsin Center for Education Research.

Wilde, J. (2009a). NAEP, the “Nation’s Report Card,” and ELL students–mathematics. *AccELLerate! The quarterly newsletter of the National Clearinghouse for English Language Acquisition, 2*(1), 2–6. Retrieved November 1, 2009, from http://www.ncela.gwu.edu/ files/accellerate/7/Accellerate_2_1.pdf

Wilde, J. (2009b). NAEP, the “Nation’s Report Card,” and ELL students. *AccELLerate! The quarterly newsletter of the National Clearinghouse for English Language Acquisition, 1*(4), 21–24. Retrieved November 1, 2009, from http://www.ncela.gwu.edu/files/ accellerate/6/Accellerate _1_4.pdf

Yoko, M. (2007). High-stakes test performance of limited English proficient students in Ohio. *Dissertation Abstracts International, 67,* n.p. A. (AAT No. 3241764)

Zwiers, J. (2006). Integrating academic language, thinking, and content: Learning scaffolds for non-native speakers in the middle grades. *Journal of English for Academic Purposes, 5,* 317–332.

Zwiers, J. (2007). Teacher practices and perspectives for developing academic language. *International Journal of Applied Linguistics, 17,* 93–116.

Appendix A

Distribution of English Proficiency (ACCESS) Levels as a Percentage of Current and Former ELL Samples

	English language skills									
Proficiency	Speaking		Listening		Reading		Writing		Composite	
Levels[a]	Cur.	Form.	Cur.	Form.	Cur.	Form.	Cur.	Form.	Cur.	Form.
Level 1	19	1	7	10	17	11	12	11	6	–[b]
Level 2	6	–	14	–	25	–	26	–	24	–
Level 3	5	–	22	2	35	3	54	29	34	–
Level 4	10	3	21	3	8	7	8	58	34	14
Level 5	11	3	28	17	14	27	–	3	2	86
Level 6	49	93	9	69	2	53	–	–	–	–
N	529	119	531	119	530	119	528	119	526	106
Median	5.2	6.0	4.1	6.0	2.9	5.9	3.0	4.0	3.6	5.2

Note. Cur. = current ELL. Form. = Former ELL. Percents do not sum up to 100 due to rounding. Only ACCESS-based proficiency level data are reported here because LAS scores for students reclassified prior to 2007 were not available; numbers of students tested within each ACCESS category (speaking, listening, reading, writing, composite) vary due to missing data.

[a]In common terminology, English proficiency levels are interpreted as follows: Level 1 = Beginning/Preproduction; Level 2 = Beginning/Production; Level 3 = Intermediate; Level 4 = Advanced Intermediate; Level 5 = Advanced; and Level 6 = Fully-English Proficient.

[b]Indicates that there were no students scoring in this category.

Appendix B

Descriptive Statistics for the Categorical and Continuous Independent Level 1 Variables Disaggregated by Language Proficiency Group

	NES (N = 17,470)		Current ELL (N = 558)		Former ELL (N = 500)	
Categorical variables	*n*	%	*n*	%	*n*	%
Female	8,544	49	233	42	254	51
Free/reduced-price lunch	9,720	56	517	93	369	74
Race						
White	9,991	57	75	13	142	28
Black	6,601	38	125	22	30	6
Hispanic	222	1	281	50	221	44
Asian/Pacific Islander	243	1	52	9	90	18
Other	390	2	25	5	17	3
Continuous variables	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age	13.62	1.01	13.99	1.18	13.49	0.94
LOR[a]			4.72	2.37	6.63	1.84

Note. Percents do not add up to 100 due to rounding.
[a]LOR = length of residence, measured as time since first U.S. school enrollment starting in kindergarten; 29 former ELLs did not have LOR data.

Appendix C

Student Academic Proficiency (KCCT) Level in Reading and Mathematics Disaggregated by Language Proficiency Group

	NES ($N = 17{,}470$)		Current ELL ($N = 558$)		Former ELL ($N = 500$)	
Proficiency levels	*n*	%	*n*	%	*n*	%
KCCT READ						
Novice	1,631	9	190	34	10	2
Apprentice	5,093	29	239	43	97	19
Proficient	8,517	49	122	22	297	59
Distinguished	2,224	13	7	1	96	19
KCCT MATH						
Novice	3,641	21	267	48	41	8
Apprentice	4,942	28	208	37	131	26
Proficient	5,720	33	78	14	223	45
Distinguished	3,162	18	5	1	105	21

Note. Percents do not sum up to 100 due to rounding.