American Educational Research Journal
August 2020, Vol. 57, No. 4, pp. 1791–1836
DOI: 10.3102/0002831219887137

Elementary English Learner Classroom Composition and Academic Achievement: The Role of Classroom-Level Segregation, Number of English Proficiency Levels, and Opportunity to Learn

Peggy Estrada
University of California, Santa Cruz
Haiwen Wang
SRI International
Timea Farkas
University of California, Santa Cruz

Using mixed methods, we investigated (a) the association of the extent of English learner (EL) classroom-level segregation (proportion EL) and number of EL English proficiency levels with elementary EL academic achievement, using 2 years of administrative data, and (b) school staff–reported opportunity to learn–related advantages and disadvantages in segregated versus

---

Peggy Estrada, PhD, is an associate research scientist in the Latin American and Latino Studies Department, Merrill Faculty Services, University of California, 1156 High Street, Santa Cruz, CA 95064; e-mail: *peggye@ucsc.edu*. She has two primary areas of research: linguistic, instructional, and curricular placement policies and practices for students with a primary language other than English in the home and their consequences for core curricular access, achievement, and integration, and pedagogy, instructional quality, and achievement in culturally and linguistically diverse classrooms, with a special focus on Latino students. In partnership with districts, she answers field-, policy-, and practice-relevant questions using a sociocultural lens and mixed methods.

Haiwen Wang, PhD, is a senior researcher at SRI International, 333 Ravenswood Avenue, BS356, Menlo Park, CA 94025; e-mail: *haiwen.wang@sri.com*. She has more than 15 years of experience in education research and expertise in quantitative research design, statistical modeling, and applying rigorous research methodology in education research.

Timea Farkas, PhD, is an assistant research scientist in the Latin American and Latino Studies Department, Merrill Faculty Services, University of California, 1156 High Street, Santa Cruz, CA 95064; e-mail: *tfarkas1@ucsc.edu*. She investigates K–12 English learner education, and traditional gender ideologies and their relation to outcomes in a variety of domains, including education, future possible selves, and close relationships.

integrated compositions, using 3 years of interviews. Findings were corroborative across methods. After accounting for student-, classroom-, and school-level covariates, we found that ELs in more segregated classrooms exhibited lower performance, on average, on state tests of English language arts, mathematics, and English proficiency, and little evidence that classroom number of EL English proficiency levels was related to achievement. School staff consistently detailed the instructional, academic, and socio-emotional opportunities to learn afforded by the diversity/heterogeneity of integrated classrooms.

KEYWORDS: achievement, classroom composition, English learner, segregation, opportunity to learn

One way in which school districts attempt to provide English learner students (ELs) legally mandated language services, improve their performance, and meet accountabilities is through classroom composition policies that prescribe the mix of students by language status (e.g., EL, English-only speaker [EO]) and by EL English proficiency (ELP) levels (beginning through advanced). Across the nation, the vast majority of ELs receive English-only instruction (Sugarman, 2018). Reflecting assumptions about the benefits of different classroom compositions for such instruction, variation in the extent of EL classroom-level segregation exists across states, districts, and schools. In California, a local control state where English-only instruction[1] for ELs dominates, districts choose whether to implement classroom-level EL segregation or integration with non-EL peers districtwide, or they permit schools to decide (e.g., San Diego Unified School District, 2005; San Jose Unified School District, 2018). The notion that language proficiency is the primary challenge to EL school success can dominate policies, obscuring a confluence of converging sociocultural and sociopolitical factors pertinent to school success.

Central among these sociocultural and sociopolitical factors is the deficit EL definition (i.e., failing an ELP test; Estrada, Park, & Farkas, 2018). In a society valuing monolingual English fluency (e.g., Pennycook, 2002, 2007), EL is a deviance label (Estrada et al., 2018; Thompson, 2015; Umansky, 2016). The meager definition obscures the competencies ELs bring to school, including the potential for multilingualism. Unfortunately, it also confounds language status with performance because to reclassify as fluent English proficient (RFEP), ELs must pass state ELP tests and, in 28 states, additional content standards performance criteria (Estrada & Wang, 2018; Wolf et al., 2008). As a result, what is often overlooked is the fact that the lower average EL versus non-EL performance is in great part an artifact of policies historically excluding from the EL group former ELs who have successfully reclassified (Saunders & Marcelletti, 2013). Easily lost from view as well are powerful socioeconomic forces that tip the playing field negatively for the majority of ELs and may pose greater obstacles to school success. Among these are

poverty; limited social capital, including lower levels of parent education; residing in segregated, linguistically isolated communities; and attending underresourced schools (EdSource, 2008; Flores, Batalova, & Fix, 2012; Gándara, Rumberger, Maxwell-Jolly, & Callahan, 2003; National Academies of Sciences, Engineering, and Medicine [NASEM], 2017). These factors, coupled with legal mandates for language services and accountability pressures, can give rise to policymakers and practitioners perceiving ELs through a "language deficit" lens and to policies that segregate ELs for "specialized" instruction.

Within English-only instructional programs, classroom-level EL segregation, while ostensibly bringing together students with similar language needs, simultaneously concentrates the myriad barriers to success in classrooms filled with students whose continued EL language status signifies concurrently not reclassifying and not meeting performance standards. Yet there is a dearth of EL-specific research regarding whether more or less segregated classroom compositions, or those with fewer or greater numbers of ELP levels, are associated with elementary students' opportunities to learn (OTL) or their achievement outcomes (Goldenberg, 2013; Saunders & Goldenberg, 2010). With little evidence for guiding policy, administrators often rely on personal experiences and perspectives. To begin to fill this gap, employing mixed methods, we (a) describe the association of the extent of EL classroom-level segregation (proportion EL) and number of ELP levels with EL elementary academic achievement, using 2 years of cross-sectional administrative data; (b) describe the school staff–reported OTL-related advantages and disadvantages in segregated EL versus integrated EL and non-EL peer classroom compositions, using 3 years of interviews; (c) examine the congruence of findings with theory and across methods; (d) develop theory about the possible processes involved; (e) identify policy implications; and (f) pinpoint further hypotheses.

Theoretical Framework

Sociocultural, second language acquisition (SLA), and labeling theories suggest that classroom composition plays a key role in EL educational outcomes by affording or constraining OTL, language and literacy development, and stigmatization. From both sociocultural and SLA perspectives, English-speaking and higher performing peers afford OTL. Sociocultural theory situates individual development and learning in the social and cultural context—within shared activities, mediated by language (Rogoff, 1990; Tharp, Estrada, Dalton, & Yamauchi, 2000; Tharp & Gallimore, 1988; Vygotsky, 1978). In school, effective teaching and learning involve teachers shouldering moderate to major levels of assisted performance in the zone of proximal development[2] *and* harnessing peer interaction for learning by creating activities and groups to use individual competencies complementarily (Tharp & Gallimore, 1988).

The primacy of language in literacy and content learning for all learners is well understood from both sociocultural and SLA perspectives (August & Shanahan, 2006; NASEM, 2017; Saunders, Goldenberg, & Marcelletti, 2013; Tharp & Gallimore, 1988). Language is central because learning requires understanding the abstract language used in school and using it as a tool for acquiring new knowledge (Carroll, 1972). Indeed, meaningful discourse is both the vehicle and the goal for developing content discourse competencies and literacy, word meanings, and conceptual structures (Tharp & Gallimore, 1988). For ELs, successfully participating in such discourse and concept development requires attaining proficiency in the abstract English used in school reading and writing. Limited English proficiency makes decoding print challenging and comprehension difficult due to limited understanding of word meanings and the ways they modify and relate to one another (August & Shannahan, 2006; Tharp & Gallimore, 1988).

From both cognitivist and language socialization SLA perspectives,[3] the key to attaining such English proficiency is providing ELs opportunities for collaborating and interacting with English-proficient peers (e.g., Atkinson, 2011; Duff & Talmy, 2011; Ellis, 2005; NASEM, 2017). Such opportunities are intrinsically motivating and crucial for producing pragmatic meaning—a vehicle for activating existing and creating additional linguistic resources—which is necessary for fluent second language acquisition (DeKeyser, 1998; Ellis, 2005). They also afford extensive second language (L2) input and opportunities for producing sustained, stretched output on oral and written tasks. Finally, interacting and collaborating with target language–proficient peers involve negotiating meaning via corrective feedback and modifying output, thus creating new language resources and fostering L2 acquisition (Ellis, 2005; Long, 1996). Heterogeneous, integrated EL classroom compositions afford instructional moves that provide ELs such opportunities. Language socialization approaches to SLA stress that conducive social, political, and cultural circumstances are necessary for such opportunities to foster L2 acquisition. That is, access and participation depend on teachers and peers positioning learners favorably as legitimate, engaged members of a language community (Duff & Talmy, 2011).

In contrast, labeling theory, which argues that socially constructed deficit labels such as EL incur costs, delineates the processes by which classroom-level EL segregation may foster stigmatization, thus positioning ELs unfavorably in their school language communities. Stigmatization occurs when labeling, stereotyping, separation of the labeled into distinct categories, and status loss converge with institutional power that permits the full execution of disapproval, rejection, exclusion, and discrimination to unfold (Link & Phelan, 2001). Although such labels are culturally situated and typically oversimplified, once we label differences, we take their validity for granted. Labels signifying undesirable characteristics affect perceptions, leading to stereotyping, which becomes the rationale for separating *us* from *them*. Status loss occurs

when labeled persons are placed low in a status hierarchy, rationalized by their undesirable characteristics. Discrimination need not be intentional; instead, it can involve institutional practices—such as curricular placement segregation—that disadvantage the labeled group. Related research indicates that once constructed, categories are instantiated in institutional representations of ELs and re-created in classrooms via teacher-student interactions and societal representations of languages and ethnicities (Harklau, 2000). As played out in classroom teaching and learning behaviors, these representations of ELs are likely to affect teacher perceptions and expectations and ELs' school experiences, including OTL (Estrada et al., 2018; McDonnell, 1995). Thus, the deficit EL label combined with policies that segregate ELs are precisely the kinds of conditions that foster stigmatization, which is related to poor educational and psychological outcomes (Estrada, 2018; Estrada et al., 2018; Link & Phelan, 2001).

Our conceptual framework indicates that in schools, teacher-student and peer interactions and relationships constitute the crucial social plane in which students' minds are created, exerting their influence primarily through assisted performance and the development of intersubjectivity, made possible by joint activity and meaningful discourse (Tharp et al., 2000). Classroom composition, then, is a powerful, wide-reaching policy lever precisely because it largely determines the pool of eligible teachers and peers for instruction, learning, language interactions, and relationships (see Tharp et al., 2000). Thus, who is in the classroom matters, and more or less segregated compositions may afford students different OTL. In turn, OTL may underlie any observed associations of classroom composition and achievement. For example, integrated classrooms may afford greater OTL due to the presence of English-proficient peers—if teachers use the affordance for eliciting language production in structured activities in which ELs and English speakers interact and collaborate. Similarly, they may also foster more positive teacher perceptions and expectations and more rigorous instruction (Estrada et al., 2018). In contrast, classroom-level segregation based on the EL label denies ELs English-proficient, higher performing peer resources for collaboration and discourse, English language learning, and social capital (Burton & Welsh, 2015) and promotes stigmatization (e.g., Estrada et al., 2018).

Literature Review

Consistent with our theoretical framework, the few elementary EL-specific studies demonstrate negative relations between classroom-level segregation and EL outcomes and point to diminished OTL as an underlying factor. Gándara and colleagues (2003) reported that in segregated classrooms lack of exposure to higher achieving classmates can hinder EL language development. Ethnographic, interview, and survey research has shown that separating ELs from non-EL peers resulted in isolation and less rigorous

content instruction for ELs (Gándara & Orfield, 2012; Lillie et al., 2010; Ríos-Aguilar, González-Canche, & Moll, 2012). A single study examining classroom composition, OTL, and achievement found that in a sample of Grade 8 math classrooms, (a) higher concentration of ELs was associated with lower content coverage and lower math scores, (b) higher content coverage predicted higher math achievement, and (c) ELs reported receiving lower content coverage compared with non-ELs (Abedi & Herman, 2010). These studies suggest that classroom composition is associated with OTL and both classroom composition and OTL may predict EL achievement.

These few EL-specific findings are consistent with elementary-level general population research indicating that segregating students by achievement at the classroom level reduces OTL and depresses outcomes for lower achieving students, with little or no benefit for higher achieving students (e.g., Slavin, 1987, 1989). In contrast, higher performing peers positively and cumulatively influence individual student achievement (e.g., Gottfried, 2014; Hanushek, Kain, Markman, & Rivkin, 2003; Hoxby, 2000), and they are associated with positive student socio-emotional outcomes (Epstein, 1983; Ryan & Deci, 2000). Because ELs are defined by not meeting ELP and, often, content achievement criteria, policies that separate them can result in de facto classroom-level ability grouping of the sort deleterious to lower achieving students.

To our knowledge, no research has examined the question of number of ELP levels in classrooms. Saunders and Goldenberg (2010) concluded that no EL-specific research directly addressed questions regarding grouping ELs at the classroom level or for English language development (ELD) instruction specifically. Based on general population ability-grouping research and sparse research on ELD instruction, they recommended that schools enhance EL achievement by (a) placing ELs in mixed-ability and mixed language status, heterogeneous classrooms and (b) grouping them within classrooms (and possibly across classrooms) by ELP level for designated ELD during a separate time block (Saunders et al., 2013). They emphasized that ELs should not be segregated into classrooms consisting of only ELs, "much less into classrooms consisting of all low-achieving ELs" (p. 57).

Consistent with sociocultural and SLA theories' emphasis on meaningful discourse within the social plane as the lynchpin in language and content learning, researchers increasingly agree that ELs need frequent, structured academic teacher-student and peer interaction with proficient English speakers (e.g., Fillmore & Snow, 2000; Gersten et al., 2007; Greenfader, Brouillette, & Farkas, 2015; NASEM, 2017). The limited extant research demonstrates that such interaction and collaboration are associated with positive EL academic outcomes (NASEM, 2017). One study found that compared with controls, K–2 ELs who participated in an ELD intervention that included peer verbal interactions made greater gains in ELP (Greenfader et al., 2015.) In another study, higher level teacher questioning and meaning-focused peer interaction with and about written text were related to better Grades 3–5 reading

comprehension, especially for ELs (O'Day, 2009). Similarly, Calhoon, Al Otaiba, Cihak, King, and Avalos (2007) found that a combination of teacher-directed, code-focused instruction and peer pair work focused on code, reading, and comprehension was related to higher Grade 1 reading scores. Finally, Zhang, Anderson, and Nguyen-Jahiel (2013) showed that peer-led collaborative reasoning discussions among EL and non-EL peers increased comprehension and enhanced ELs' interest, engagement, and English-learning attitudes. Across studies, researchers point to the role of accurate feedback, a focus on language and literacy development, and structured discourse opportunities with higher performing, more proficient speakers in contexts fostering high engagement and motivation.

Consistent with labeling theory, the available research, primarily at the secondary level, shows that the deficit EL label, institutional representations highlighting "deficits" needing remedying, and curricular placement policies that segregate ELs are associated with classroom-level instantiations of these representations and EL stigmatization. A case study of secondary ELs showed that the negative connotations of the long-term EL label often blinded school staff to students' successes, abilities, and experiences (Thompson, 2015). A quasi-experiment found that the EL label negatively affected performance on state tests of English language arts (ELA) and math (Umansky, 2016).

Similarly, Harklau (2000) found that when a high school's institutional representation of ELs was "well-behaved, hardworking, persistent immigrants," teachers considered ELs "good kids." Teacher and students' interactions re-created positive identity in the context of assignments that cast ELs in a favorable light. Students found such assignments motivating, delivered, and behaved well. Key also was that EL identity formation occurred primarily within mainstream classroom interactions because ELs were separated only for their English for Speakers of Other Languages class. In contrast, in a community college with deficit representations of the same ELs, teachers adopted deficient views and stigmatized ELs, which was reflected in demotivating assignments. Students resisted in response. Thus, positive institutional representation engendered teaching that increased OTL by engaging students in learning; the latter fostered the opposite.

Separate secondary EL Curricular Streams are associated with negative teacher conceptions of ELs and institutional and teacher behaviors that reduce OTL (Estrada, 2014, 2018; Estrada et al., 2018). Teachers in one school in Estrada's (2014) study reported that separating students into majority-EL, multirostered sheltered content courses resulted in limited opportunities for participation in the full curriculum and isolated and stigmatized both ELs and their teachers. Teachers' peers avoided teaching these courses; student peers viewed ELs as less able. Teachers (Dabach, 2014; Estrada, 2018; Estrada et al., 2018) and students (Thompson, 2015) have also reported that EL sheltered courses are stigmatized spaces, perceived as "easier" and signifying less intelligence to both students and peers.

The available evidence leaves many unknowns, including how the extent of classroom-level EL segregation and number of ELP levels are associated with EL academic outcomes and whether these aspects of classroom composition afford different OTL at the elementary school level. This study begins to fill these gaps. Combining theory and the sparse evidence, we hypothesize that greater classroom-level EL segregation (higher proportion EL) will be negatively associated with EL outcomes. In the absence of specific research, our investigation regarding the number of ELP levels within classrooms is exploratory. Teachers of ELs have to provide both content instruction and ELD instruction, adding to the usual range of performance capacities teachers must respond to. Too many ELP levels within classrooms could pose additional challenges for organizing and tailoring designated ELD, integrated ELD, and content instruction. We hypothesize that ELs in classrooms with four to five levels of ELP may exhibit lower levels of academic performance.

Research and District Policy Context

Part of a large-scale, 4-year longitudinal project, the present study occurred in District 1, a high-poverty urban district. Compared with the state, it had more free and reduced-price lunch eligible students (77% vs. 58%), Latinos (73% vs. 52%), ELs (27% vs. 22%), and Spanish-speaking ELs (94% vs. 85%; California Department of Education [CDE], 2012). It is experienced in internal research, developing and implementing EL educational policies and programs, and partnering with researchers.

For its English-only instructional program, which educated 89% of its ELs, District 1 had recently enforced a classroom composition policy that largely separated its elementary EL students from their non-EL peers. The natural variation in the proportion of ELs and number of ELP levels in classrooms in District 1 provided an opportunity to investigate the association between the extent of classroom-level segregation and EL outcomes. District 1 policy was as follows: For ELs at beginning to intermediate ELP levels (Levels 1–3), it directed schools to create classrooms with 100% EL language status students with two or three adjacent levels of proficiency—to the extent possible. When this composition was not feasible, schools were to backfill with ELs at early-advanced and advanced levels of ELP (Levels 4 and 5). Thus, this policy could result in EL-only classrooms with five ELP levels. According to district staff, the rationale for the policy was that the increased homogeneity in separate EL classrooms would increase teacher focus on ELs' ELA and ELD instructional needs and therefore increase instructional effectiveness and—by extension—improve student learning. The policy was in place during the study, with increased district enforcement in the final year, 2013–2014. Recognizing the paucity of EL-specific research, both District 1 staff and researchers agreed that gathering empirical evidence on the following questions was crucial.

Research Questions

Research Question 1: On average, do ELs in more segregated classrooms exhibit lower content and language performance?

Research Question 2: On average, do ELs in classrooms with four to five levels of ELP compared with those in classrooms with one to three levels exhibit lower content and language performance?

Research Question 3: What OTL-related advantages and disadvantages do school staff report in integrated versus segregated EL classroom compositions, and how do they explain them?

Research Question 4: How do the quantitative and qualitative findings corroborate or help explain each other?

Methods

The purpose of this mixed methods study was to (a) use quantitative hypothesis testing to reveal the possible academic performance consequences of EL classroom-level segregation, (b) use qualitative analyses to uncover advantages and disadvantages of segregated versus integrated EL classrooms that point to possible underlying mechanisms, and (c) examine the congruence of the findings with theory and across methods (Plano Clark & Badiee, 2010). We used a convergent, parallel mixed-methods design (Creswell & Plano Clark, 2018). For investigating Research Question 1 and Research Question 2, we used districtwide elementary K–6 administrative data (baseline and 2 years of outcomes [2013, 2014]), taking advantage of the natural variation in EL classroom composition (proportion EL and number of ELP levels). For answering Research Question 3, we used 3 years (2012, 2013, 2014) of (a) policy documents, (b) district staff interviews, and (c) school staff interviews from five randomly selected schools. We examined the congruence of the results with theory and across methods, then derived an integrated interpretation.

Quantitative Samples

Classroom Samples

The base 2012–2013 and 2013–2014 samples were composed of 5,726 Grades 2–6 classrooms and 5,311 Grades 1–6 classrooms from 468 and 437 schools, respectively. All classrooms were from K–5 or K–6 configured schools, excluding independent charters, for which no data were available. The classroom samples represented 64% and 57% of all District 1 classrooms at those grade levels, respectively, after eliminating those with (a) fewer than 10 students, (b) fewer than 5 ELs (the minimum number of ELs needed to represent the range of ELP levels on the state test of English proficiency), and (c) nonadjacent ELP levels, which were extremely rare (0.14% and 0.18% in 2012–2013 and 2013–2014, respectively). On average, classrooms

were majority EL (63% in 2012–13 and 76% in 2013–2014). The increase in the second year coincided with enforcement of the classroom composition policy (see Table 1).

Student Analytic Samples

Grades 2–6 (n = 54,176) and Grades 1–6 (n = 69,991) ELs from the classroom samples (described above) constituted the base analytic samples for two cross-sectional analyses. Few outcomes prior to Grade 2 were available; at District 1's request, we included Grade 1 in the second round of analyses. The vast majority of ELs were from Spanish-speaking homes, free or reduced-price lunch eligible, and U.S. born, and had entered U.S. schools in kindergarten or before (see Table 1). In both years, about half of them were at ELP Level 3 (intermediate), and about a fifth to a quarter were at ELP Levels 4 and 5 combined (early advanced and advanced). Close to half (40% in 2012–2013 and 51% in 2013–2014) of ELs were in classrooms with three ELP levels, and roughly 30% were in classrooms with four ELP levels.

Quantitative Measures

Each fall, District 1 provided student performance, language status, and demographic data. In general, the mobility rate for students was low. In both years, the majority of measures had no missing data in the base analytic samples. Across the two years of analyses, a few predictor variables had low levels of missing data (<5%). Outcome measures in 2012–2013 had slightly higher rates (6% to 8%) of missing data but very low rates (4%) in 2013–2014. For analyses, we applied casewise deletion of students with missing data.

English Language Proficiency Performance

Annual California English Language Development Test (CELDT) administration occurred July through October, with scores reported in late January to early February (CDE, 2015). For the 2011–2012, 2012–2013, 2013–2014, and 2014–2015 academic years, District 1 provided individual student overall CELDT scale scores and ELP levels (1 = beginning, 2 = early intermediate, 3 = intermediate, 4 = early advanced, 5 = advanced). Because CELDT test scores are not directly comparable across grade levels, we standardized them at the grade-level mean and combined grade levels for the whole-group analyses.

English Language Arts and Math Performance

Through 2012–2013, California Standards Test (CST) administration occurred in spring, with scores reported in August. CST scale scores ranged from 150 to 600 (300–349 = basic, 350 = proficient). The state dropped, and did not replace, CST testing in 2013–2014, precluding the use of those scores in that year's analysis. Because CST test scores are not directly comparable

Table 1

Covariate and Outcome Descriptive Statistics

Variable	2012–2013	2013–2014
Student level (%)		
Female	47.51	47.83
Hispanic	94.17	93.92
Asian	2.69	2.91
FRLP	83.27	87.08
Spanish home language	94.48	94.34
Entered K	95.94	—
U.S. born	91.68	93.76
ELP Level 1	4.37	12.22
ELP Level 2	15.54	20.53
ELP Level 3	51.93	45.85
ELP Level 4	23.10	17.13
ELP Level 5	5.05	4.27
Grade 1		26.27
Grade 2	37.10	27.62
Grade 3	24.84	18.85
Grade 4	19.41	14.69
Grade 5	16.29	11.04
Grade 6	2.36	1.53
Classroom level (%)		
One ELP level	2.27	1.38
Two ELP levels	17.58	13.45
Three ELP levels	40.40	50.79
Four ELP levels	30.14	27.15
Five ELP levels	9.61	7.23

(continued)

Table 1 **(continued)**

Variable	2012–2013				2013–2014			
	M	*SD*	Range (Actual)	Range (Possible)	*M*	*SD*	Range (Actual)	Range (Possible)
Prop. EL	0.63	0.26	0.11–1.00	0–1.00	0.76	0.27	0.11–1.00	0–1.00
School level (%)								
Title 1	96.45				96.95			
School level								
Total enrollment	729.06	236.64	135–2,038		705.01	233.22	120–1,982	
Prop. EL	0.44	0.12	0.01–0.78	0–1.00	0.48	0.13	0.02–0.82	0–1.00
Prop. RFEP	0.28	0.08	0.01–0.55	0–1.00	0.21	0.06	0.01–0.57	0–1.00
Prop. FRLP	0.87	0.12	0.04–1.00	0–1.00	0.86	0.10	0.04–0.98	0–1.00
Prop. Hispanic	0.84	0.17	0.03–1.00	0–1.00	0.85	0.17	0.03–1.00	0–1.00
API	774.73	50.58	543–984	200–1,000	767.53	53.76	530–981	200–1,000
EL API	752.04	48.70	501–985	200–1,000	741.50	53.26	462–979	200–1,000
Similar schools rank	5.64	2.74	1–10	1–10	—	—	—	—
Outcome[a]								
Reclassification to FEP	0.29		0–1.00	0–1.00	0.23		0–1.00	0–1.00

Note. FRLP = free or reduced-price lunch eligible; Entered K = entered U.S. schools in kindergarten or prior; ELP = English language proficiency; Prop. = proportion; EL = English learner; FEP = fluent English proficient; RFEP = reclassified fluent English proficient; API = Academic Performance Index; *SD* = standard deviation.

[a]Because California Standards Test and California English Language Development Test scores are not directly comparable across grade levels, we present descriptive statistics for these outcomes by grade level in Table 3.

across grade levels, we standardized them at the grade-level mean and combined grade levels for the whole-group analyses.

Reclassification

Administrative data indicated whether each student reclassified as FEP (1 = *reclassified as FEP*, 0 = *remained EL*) in the years 2012–2013 and 2013–2014.

Classroom Proportion EL and Number of ELP Levels

Administrative enrollment data indicated individual student language status (e.g., EL, EO) and ELP level. Using these data, for each classroom we (a) operationalized the extent of segregation as the proportion of ELs and (b) calculated the number of adjacent ELP levels.

Other Student-Level Covariates

Student-level covariates included gender, ethnicity, free or reduced-price lunch eligibility, home language, country of birth, whether year of U.S. school entry was kindergarten or before, grade level (a series of dummy variables, with the lowest grade level functioning as the omitted comparison group for each analysis: for the first year of analyses, Grade 2 for models predicting CELDT and Grade 3 for models predicting CST and reclassification; for the second year of analyses, Grade 1 for models predicting CELDT and Grade 3 for models predicting reclassification), previous year's ELP level, overall CELDT score, CST-ELA score, and CST-math score. Whether year of entry was kindergarten or before was excluded from the 2013–2014 analysis due to considerable missing data.

Other Classroom-Level Covariates

Our theoretical framework and concerns about overadjustment bias guided the decision to include our main predictors of interest—classroom proportion EL and classroom number of ELP levels—but not include other classroom-level covariates. Our theoretical framework indicates that classroom-level factors such as mean peer performance and demographics (a probable proxy for social capital) are on the causal pathway between EL classroom composition and EL outcomes. Therefore, including such variables in our models would lead to overadjustment bias (Schisterman, Cole, & Platt, 2009).

Other School-Level Covariates

School-level covariates included total student enrollment, Title 1 status, Academic Performance Index (API; a measure based primarily on state standards tests results, with a target of 800 or higher for all schools; CDE, 2019), EL API, and similar schools rank (a 1–10 ranking of overall school

performance compared with 100 demographically similar schools statewide; CDE, 2019). Similar schools rank was excluded from the 2013–2014 analysis due to considerable missing data. Additional school-level variables included the proportion of enrolled students who were ELs, reclassified FEP, free or reduced-price lunch eligible, or Hispanic.

Quantitative Data Analysis

To examine the relation of classroom-level proportion EL and number of ELP levels with EL outcomes, we estimated three-level hierarchical models with student-, classroom-, and school-level covariates. The continuous measure of classroom-level proportion EL allowed us to detect nuance in any observed relation, such as whether it was linear or curvilinear, while maximizing the number of classrooms in the analytic sample. Using one to three classroom ELP levels as the comparison (vs. four to five ELP levels) allowed both testing our hypothesis that in classrooms with four to five ELP levels ELs may exhibit lower performance as well as the utility of the policy, which the district was keen for us to do. Using maximum likelihood estimation, we predicted language and content outcomes with classroom proportion EL and number of classroom ELP levels, along with other covariates. Separate models predicting CST-ELA scores, CST-math scores, and reclassification included prior-year baseline performance and concurrent-year covariates, including classroom proportion EL and number of ELP levels. Due to the timing of CELDT administration (prior to or at the beginning of each academic year), models predicting CELDT scores included prior-year baseline performance and covariates, including classroom proportion EL and number of ELP levels. For this reason, the corresponding CELDT outcome scores were from 2013–2014 and 2014–2015. Note that we conducted the analyses predicting CST-ELA, CST-math, and reclassification on Grades 3–6, including baseline CST scores as covariates, because CST testing began in Grade 2. We conducted these analyses separately for Grade 2 without including baseline CST scores as covariates. All continuous predictors were grand-mean centered. The three-level model was specified as follows:

$$\text{Student level}: Y_{ijk} = \pi_{0jk} + \pi_{1jk} X_{ijk} + e_{ijk}$$

$$\text{Classroom level}: \pi_{0jk} = \beta_{00k} + \beta_{01k} Z_{jk} + r_{0jk}$$

$$\text{School level}: \beta_{00k} = \gamma_{000} + \gamma_{001} W_k + \mu_{00k}$$

Y_{ijk} represents the outcome for the *i*th student in the *j*th classroom in the *k*th school. A linear model is posited for all outcomes except the dichotomous outcome of reclassification, for which we used a logit link function.

X_{ijk}, Z_{jk}, and W_k represent vectors for student-, classroom-, and school-level covariates, respectively; π_{1jk}, β_{01k}, and γ_{001} represent vectors of the coefficients for student-, classroom-, and school-level covariates, respectively. The key predictors, classroom proportion EL and number of ELP levels, were included at the classroom level. Classroom number of ELP levels was included as a dichotomous indicator, with one to three ELP levels as the reference category.

Finally, e_{ijk}, r_{0jk}, and μ_{00k} are random errors at the student, classroom, and school levels, respectively.

We conducted these analyses for the whole sample. We also conducted exploratory analyses by individual student CELDT level and by grade level. Because the results of the models by individual CELDT level (see Supplemental Tables S1–S4 in the online version of the journal) and by grade level (available upon request from the first author) were similar to those of the whole-sample models, we report only the latter. To explore whether the number of ELP levels and the proportion of ELs in classrooms interact to predict outcomes, we added interaction terms between the number of ELP levels and proportion of ELs at the classroom level. Finally, to explore whether the relation between extent of classroom segregation and outcomes was curvilinear rather than linear, we added a quadratic term for the proportion of ELs variable to the models.

Qualitative Samples

School Samples for Interviews

To study policies and practices where most ELs were enrolled and where EL subgroup outcome data were available, we restricted the pool to schools with greater than 10% ELs and more than 100 ELs. This strategy yielded a target pool containing 94% of District 1 ELs. We excluded independent charter schools due to lack of data availability. To select five District 1 elementary schools, we stratified by reclassification rate (high/low) and EL API (high/low). We defined the cutoff for *high* as the highest of three medians: the state, District 1, and District 2 (part of the larger project but not this study); similarly, the cutoff for *low* was the lowest of the three medians (available from the first author). This approach avoided a crossover effect in high/low categorization (e.g., high in one district is low in the other). We drew random samples for each cell. All the five invited elementary schools in District 1 participated across the 3 years. All five schools provided English-only instruction to ELs.

Staff Samples for Interviews

Across years, we interviewed seven district administrators and program coordinators who were experienced with ELs (10–19 years) and involved in

Table 2
School Staff Interview Samples by Year

Staff	2011–2012	2012–2013	2013–2014	No. of Unique Participants	Total Interviews
Total staff	15	14	34	43	63
District administrators and program coordinators	4	4	3	7	11
School administrators	2	2	4	5	8
Support staff	2	2	6	6	10
Teachers	7	6	21	25	34
No. of schools	2	2	5	5	

Note. We endeavored to interview the same staff over the 3 years, though it was not always possible. For this reason, total participants and total interviews are different.

EL policy development and implementation and/or accountability (see Table 2 for staff samples and interviews across years). They tended to be recent to their positions (1–2 years), corresponding with the implementation of a new EL Master Plan. We interviewed 25 teachers and 11 nonteaching staff at five schools across years. On average, school staff had been in their positions 9 to 13 years and had worked with ELs for 14 to 18 years (detailed staff demographics are available from the first author). We adopted District 1's protocols for contacting principals, soliciting participation, and arranging site visits. At each school, we interviewed 5 to 6 staff members: the principal or academic vice principal, the English language coordinator, 3 to 4 teachers, and occasionally instructional coaches and Title 3 coordinators. The district permitted direct contact with teachers only during the interview. Therefore, principals or English language coordinators facilitated site visits and selected teachers for interviews based on the characteristics we were targeting: Elementary teachers typically taught multiple subjects and ELD to Grades 3–5 ELs (across years, 3 did not teach ELD, and 5 taught Grade 2 or lower). For continuity, we endeavored to interview the same staff across years, although it was not always possible.

Qualitative Data: EL Policy Documents and Staff Interviews

Annually, to gain contextual policy and implementation knowledge, we began by applying a master set of queries to the EL Master Plan and policy documents pertaining to elementary school EL curricular placement and classroom composition. Using that knowledge, we tailored interviews for district and school staff in different roles (e.g., administrator, teacher). A common set of analytic categories, such as the rationale for the EL classroom composition policy and OTL-related advantages and disadvantages of inclusion and separation for EL student learning, drove both policy document and

interview queries. For gaining more in-depth contextual understanding of policies, implementation, and views on OTL-related and EL learning consequences, we conducted the district staff interviews first. Next, we wrote interview debriefings summarizing the responses for each analytic category and further revised school staff protocols. Piloting in nonparticipating district schools followed, along with final revisions for school staff interviews.

School staff interviews (60–90 minutes for nonteaching staff, 50 minutes for teachers) solicited (a) demographics, roles, responsibilities, and brief school descriptions and (b) site-level understanding of classroom composition policies, implementation, and views on OTL-related advantages and disadvantages of inclusion and separation for EL student learning. For example, in 2011–2012, we asked the participants about ELs' curricular placement (i.e., "Can you tell me whether ELs are scheduled in your class as a cohort?") and about the positive and negative aspects of EL placement (i.e., "What aspects of your instructional placement and Curricular Streams [Estrada, 2014] for ELs do you think work to promote academic progress and social integration, and which do you think need work?"). Knowledge gained from these staff reports spurred revisions to more specifically query the advantages and disadvantages of segregated EL versus integrated EL and non-EL classroom placements. Thus, in 2012–2013, we asked, "What are the advantages and/or disadvantages of being placed in these nearly all-EL cohorts versus clustering groups of ELs with non-EL peers (e.g., 10 in a class of 30) within a range of performance levels?" Staff responses touched on multiple domains. Again, building on that knowledge, in 2013–2014, we asked the participants to reflect on their "experience with models that separate or cluster ELs with non-ELs in classrooms" in four domains (i.e., "What are the advantages and disadvantages, if any, of each model in the following areas: ELA and ELD instruction, student learning, social, and behavioral?").

Prior to field visits each year, the first author trained the interviewers. The training focused on the semistructured interview protocol, including obtaining informed consent, listening carefully, maintaining an interested and neutral posture, when and how to use the follow-up probes, and how to restate their understandings of participants' responses and ask for feedback on whether they accurately reflected participants' intended meaning. The first author also modeled field interviewing, observed the interviewers, and provided feedback. The training also encompassed review of district EL policies and school site characteristics (e.g., demographics), the site visit protocol, and subsequent qualitative coding, analysis, and school site debriefing write-up. Across years, six researchers in total interviewed staff in the winter and spring of 2011–2012, late fall and spring of 2012–2013, and winter and spring of 2013–2014. The interviews were audiotaped and transcribed. District 1 did not permit participant compensation.

Analyses of Qualitative Data

We summarized the answers to each policy document query, analyzing them for clarity, and as noted above, we used the knowledge gained for developing interview protocols. Descriptive analysis of the interview data proceeded similarly each year as follows (except that in 2011–2012 we did it by hand and did not use qualitative software). First, we checked the accuracy of the transcriptions with the audiotapes and corrected any errors. Next, we read and coded the interviews, based on codes derived from the analytic categories (using Dedoose for 2012–2013 data and NVivo for 2013–2014 data). Each year, codes were initially derived from analytic categories, followed by an iterative process that involved (a) deliberating the meaning of the codes, (b) applying them to a set of transcripts, (c) convening to discuss and work out disagreements, (d) revising the codes and finalizing them, and (e) applying the final codes. Over the years of study, the codes reflected the increasing specificity of the interview questions, based on our accumulating knowledge. For 2013–2014, for example, the codes were Disadvantages of Separation, Disadvantages of Inclusion, Advantages of Separation, and Advantages of Inclusion for each of the following domains: instruction, student learning, social and behavioral, and planning.

Writing school debriefings followed, using a debriefing guide, which mapped the interview questions onto our analytic categories (e.g., Advantages and Disadvantages of Integration and Separation) and functioned as a template for writing a description of findings across interviews for each school. The debriefing guide provided details on (a) preparing for writing, such as building background knowledge of district policies; (b) reading coded transcript excerpts and constructing summary statements, including the level of their specificity; and (c) using evidence and quotes to back up each summary statement. We used an iterative process of reading the first interview excerpt, typically the principal's, developing a "tentative" summary statement, then reviewing, revising, and adding variations and minority views after reading the remaining pertinent nonteaching staff and teacher excerpts. Evidence for each summary statement consisted of illustrative examples and quotes. This process allowed us to identify trends, exceptions, and differences and similarities in administrators' and teachers' responses. We separated summary statements for nonteaching staff and teaching staff only when they represented different views. Finally, to summarize findings, we extracted themes, synthesizing summary statements and evidence across school debriefings for each analytic category (Braun & Clarke, 2006). We read summary statements across schools for each analytic category to capture the themes to which the weight of the evidence pointed. The themes represented a slightly higher level of summary than those in the debriefings and included theme variations and/or dimensions, which provided illustrative detail. We noted the strength of the themes,

reflecting the extent to which staff held such views (e.g., majority, some, few), and also recorded minority views and/or exceptions to the themes. Researchers debriefed biweekly, sharing emerging themes and contesting with confirming and disconfirming evidence until they reached consensus.

Results

Research Question 1: On Average, Do ELs in More Segregated Classrooms Exhibit Lower Content and Language Performance?

Descriptive statistics for baseline and outcome variables are in Table 1 (reclassified FEP) and Table 3 (CSTs and CELDT overall). Each year, roughly a quarter (2012–2013) to a third (2013–2014) of students reclassified. In both years, CST mean scores by grade level ranged from just below basic to basic in the baseline and outcome years. Mean CELDT baseline and outcome scores by grade level ranged from intermediate to early advanced.

ELs in classrooms with higher proportions of ELs exhibited lower performance, on average, on state tests of ELA and math in the single year for which data were available (see Table 4) and on the state test of ELP in both years (see Table 5), after accounting for student-, classroom-, and school-level covariates. These findings supported our hypothesis that more highly segregated classrooms would be negatively associated with EL outcomes. The proportion of ELs in classrooms was unrelated to current-year reclassification in both years (Table 6). These results may have reflected that, due to the timing of the assessments, meeting ELP and ELA content standards criteria for reclassifying in the current year were mostly a result of the prior year's classroom composition and practices.

Research Question 2: On Average, Do ELs in Classrooms With Four to Five Levels of ELP Compared With Those in Classrooms With One to Three Levels Exhibit Lower Content and Language Performance?

Our exploratory hypothesis that ELs would exhibit lower academic performance, on average, in classrooms with four to five versus one to three levels of ELP was largely unsupported. After accounting for school-, classroom-, and student-level covariates, number of classroom ELP levels was unrelated to CST-ELA and CST-math performance (see Table 4). Similarly, after accounting for covariates, number of classroom ELP levels was unrelated to 2013–2014 CELDT performance; it showed a small but statistically significant negative association with 2014–2015 CELDT performance, indicating that, on average, ELs in classrooms with four to five ELP levels performed at a slightly lower level than those in classrooms with one to three ELP levels (see Table 5). Number of classroom ELP levels was unrelated to current-year reclassification rates in both years (see Table 6).

Table 3
California Standards and English Proficiency Test Descriptive Statistics by Grade Level

	2011–2012 (Baseline)			2012–2013 (Outcome)		
Grade Level in 2012–2013	*M*	*SD*	*n*	*M*	*SD*	*n*
		CST-ELA				
2				341.06	56.91	20,025
3	325.08	47.74	13,362	303.57	46.02	12,845
4	291.73	42.09	9,953	320.76	38.53	9,591
5	316.96	33.98	8,056	308.61	34.75	7,883
6	294.08	30.02	1,126	298.52	32.59	1,085
		CST-Math				
2				355.31	75.55	20,037
3	334.08	67.28	13,368	354.26	75.22	12,917
4	345.27	68.93	10,043	346.73	64.18	9,786
5	336.06	60.35	8,221	325.26	63.86	8,044
6	311.05	58.33	1,155	296.58	43.34	1,127
	2012–2013 (Baseline)			2013–2014 (Outcome)		
Grade Level in 2012–2013	*M*	*SD*	*n*	*M*	*SD*	*n*
		CELDT				
2	487.96	47.28	20,033	489.82	43.63	12,786
3	489.86	40.02	12,838	520.22	41.02	9,639
4	516.36	37.90	10,261	542.25	39.32	7,071
5	538.80	38.42	7,969	534.42	43.60	4,204
6	537.38	43.53	1,151	551.74	42.97	844
	2013–2014 (Baseline)			2014–2015 (Outcome)		
Grade Level in 2013–2014	*M*	*SD*	*n*	*M*	*SD*	*n*
		CELDT				
1	455.53	51.09	18,363	485.37	45.94	16,378
2	489.01	46.10	19,225	490.42	46.28	13,488
3	488.58	44.39	12,581	515.03	42.35	9,616
4	519.00	42.38	10,041	540.00	41.21	7,136
5	540.23	42.43	7,111	531.56	43.41	3,806
6	534.25	45.46	988	547.36	42.72	670

Note. CST = California Standards Test; ELA = English language arts; CELDT = California English Language Development Test. CST and CELDT scores are not directly comparable across grade levels, so we standardized them at the grade-level means for all analyses. At every grade level, the range of possible CST scores is 150–600: 300 = basic (reclassification criterion cutoff), 350 = proficient. Possible Grades 1–6 CELDT scores = 184–741 (varying somewhat by grade level); criterion cutoffs for reclassifying were Grade 1 = 456, Grade 2 = 496, Grade 3 = 514, Grade 4 = 531, Grade 5 = 539, and Grade 6 = 552. The *n*s vary from the base samples because only English learners with CST or CELDT data are included. Because CST administration began at Grade 2, Grade 2 students do not have baseline CST scores.

Table 4

The Relation of Classroom Proportion of EL Students and Number of ELP Levels With EL Student CST Performance in Elementary School

	CST-ELA 2012–2013				CST-Math 2012–2013			
	Model 1		Model 2		Model 1		Model 2	
	B	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*
Intercept	0.04	0.10	0.04	0.10	–0.11	0.11	–0.13	0.11
Student level								
Baseline CST	0.61***	0.00	0.61***	0.00	0.63***	0.00	0.63***	0.00
ELP Level 1	–0.01	0.02	−0.02	0.02	0.01	0.02	0.01	0.02
ELP Level 3	0.25***	0.01	0.25***	0.01	0.16***	0.01	0.16***	0.01
ELP Level 4	0.43***	0.02	0.43***	0.02	0.27***	0.01	0.27***	0.01
ELP Level 5	0.57***	0.04	0.57***	0.04	0.40***	0.04	0.40***	0.04
Female	0.09***	0.01	0.09***	0.01	−0.01	0.01	−0.01	0.01
Hispanic	0.09	0.06	0.09	0.06	0.03	0.06	0.03	0.06
Asian	0.14**	0.04	0.14**	0.04	0.14***	0.04	0.14***	0.04
Spanish	−0.17*	0.07	−0.17*	0.07	−0.14*	0.06	−0.14*	0.06
FRPL	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Entered K	−0.19***	0.02	−0.19***	0.02	−0.15***	0.02	−0.15***	0.02
U.S. born	–0.09***	0.01	−0.09***	0.01	−0.08***	0.01	−0.08***	0.01
Grade 4	−0.01	0.02	−0.01	0.02	−0.01	0.02	−0.01	0.02
Grade 5	−0.06**	0.02	−0.06**	0.02	−0.04*	0.02	−0.04*	0.02
Grade 6	−0.07*	0.04	−0.07*	0.04	−0.02	0.04	−0.02	0.04

(continued)

Table 4 **(continued)**

	CST-ELA 2012–2013				CST-Math 2012–2013			
	Model 1		Model 2		Model 1		Model 2	
	B	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*
Classroom level								
Four to five ELP levels in class	0.00	0.02	0.00	0.02	−0.04	0.02	−0.04	0.02
Prop. EL	−0.20***	0.03	−0.20***	0.04	−0.12**	0.04	−0.12**	0.04
Four to five ELP levels × prop. EL			−0.03	0.07			0.01	0.08
Prop. EL squared			0.21	0.12			0.22	0.14
School level								
Total enrollment	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Prop. EL	0.40***	0.09	0.42***	0.09	0.46***	0.11	0.48***	0.11
Prop. RFEP	−0.22	0.12	−0.22	0.12	−0.12	0.15	−0.12	0.15
Prop. FRPL	−0.38**	0.12	−0.37**	0.12	−0.36	0.15	−0.35*	0.15
Prop. Hispanic	0.09	0.07	0.09	0.07	0.10*	0.09	0.09	0.09
API	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EL API	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00
Similar schools rank	0.00	0.01	0.00	0.01	0.00	0.00	0.01	0.01
Title 1	0.00	0.09	0.00	0.09	0.27*	0.11	0.27*	0.11

(*continued*)

Table 4 **(continued)**

	Random Parameters							
	CST-ELA 2012–2013				CST-Math 2012–2013			
	Model 1		Model 2		Model 1		Model 2	
	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*
School level								
Intercept *SD*	0.07	0.01	0.07	0.01	0.10	0.01	0.10	0.01
ICC	0.01		0.01		0.02		0.02	
Classroom level								
Intercept *SD*	0.29	0.01	0.29	0.01	0.39	0.01	0.39	0.01
ICC	0.18		0.18		0.31		0.31	
Residual *SD*	0.63	0.00	0.63	0.00	0.60	0.00	0.60	0.00
R^2	.27		.27		0.29		0.29	
n Students	29,641		29,641		30,094		30,094	
n Classrooms	3,373		3,373		3,374		3,374	
n Schools	426		426		426		426	

Note. Estimates are unstandardized. Outcome variables were standardized, thus estimates are interpretable as change in units of *SD*. *SE* = standard error; *SD* = standard deviation; CST = California Standards Test; ELA = English language arts; ELP = English language proficiency; FRPL = free or reduced-price lunch eligible; Entered K = entered U.S. schools in kindergarten or prior; Prop. = proportion; EL = English learner; RFEP = reclassified fluent English proficient; API = Academic Performance Index; ICC = intraclass correlation; R^2 = proportion of variance in the outcome explained by adding fixed factors to a model containing only the intercept and random factors.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 5

The Relation of Classroom Proportion of EL Students and Number of ELP Levels With EL Student CELDT Performance in Elementary School

	CELDT 2013–2014				CELDT 2014–2015			
	Model 1		Model 2		Model 1		Model 2	
	B	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*
Intercept	0.17	0.11	0.18	0.11	0.24***	0.06	0.27***	0.06
Student level								
Baseline CELDT	0.62***	0.01	0.62***	0.01	0.59***	0.00	0.59***	0.00
ELP Level 1	−0.12***	0.02	−0.12***	0.02	−0.03**	0.01	−0.03**	0.01
ELP Level 3	0.31***	0.01	0.31***	0.01	0.22***	0.01	0.22***	0.01
ELP Level 4	0.50***	0.02	0.50***	0.02	0.41***	0.01	0.41***	0.01
ELP Level 5	0.63***	0.03	0.63***	0.03	0.61***	0.02	0.61***	0.02
Female	0.04***	0.01	0.04***	0.01	0.07***	0.01	0.07***	0.01
Hispanic	0.00	0.06	0.00	0.06	−0.05	0.04	−0.06	0.04
Asian	0.11*	0.05	0.11*	0.05	0.16***	0.03	0.16***	0.03
Spanish home language	−0.06	0.07	−0.06	0.07	−0.09	0.04	−0.09	0.04
FRPL	−0.02	0.01	−0.02	0.01	−0.04***	0.01	−0.04***	0.01
Entered K	−0.25***	0.02	−0.25***	0.02				
U.S. born	−0.05**	0.02	−0.05***	0.02	−0.14***	0.01	−0.14***	0.01
Grade 2					0.02	0.02	0.02	0.02
Grade 3	−0.03	0.02	−0.03	0.02	0.06**	0.02	0.06**	0.02
Grade 4	−0.01	0.02	−0.01	0.02	0.07**	0.02	0.07***	0.02
Grade 5	0.02	0.02	0.02	0.02	0.08***	0.02	0.08***	0.02
Grade 6	−0.18***	0.04	−0.18***	0.04	−0.06	0.05	−0.05	0.05

(continued)

Table 5 **(continued)**

	CELDT 2013–2014				CELDT 2014–2015			
	Model 1		Model 2		Model 1		Model 2	
	B	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*
Classroom level								
Four to five ELP levels in class	−0.02	0.02	−0.02	0.02	−0.04**	0.01	−0.05**	0.01
Prop. EL	−0.15***	0.04	−0.16***	0.04	−0.14***	0.03	−0.22***	0.04
Four to five ELP levels × prop. EL			0.04	0.06			0.05	0.05
Prop. EL squared			−0.15	0.13			−0.29**	0.10
School level								
Total enrollment	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Prop. EL	0.34***	0.12	0.32**	0.12	0.24**	0.08	0.23**	0.08
Prop. RFEP	−0.11	0.16	−0.11	0.16	0.02	0.14	0.00	0.14
Prop. FRPL	−0.40*	0.16	−0.40*	0.16	−0.18	0.10	−0.18	0.10
Prop. Hispanic	0.16	0.10	0.16	0.10	0.14*	0.06	0.14*	0.06
API	0.00	0.00	0.00	0.00	0.00**	0.00	0.00**	0.00
EL API	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Title 1	0.20*	0.10	0.19	0.10	0.02	0.05	0.01	0.05
Similar schools rank	0.00	0.01	0.00	0.01				

(*continued*)

Table 5 **(continued)**

	Random Parameters							
	CELDT 2013–2014				CELDT 2014–2015			
	Model 1		Model 2		Model 1		Model 2	
	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*
School level								
Intercept *SD*	0.16	0.01	0.16	0.01	0.09	0.01	0.09	0.01
ICC	0.04		0.04		0.01		0.01	
Classroom level								
Intercept *SD*	0.31	0.01	0.31	0.01	0.32	0.01	0.32	0.01
ICC	0.19		0.19		0.18		0.18	
Residual *SD*	0.72	0.00	0.72	0.00	0.70	0.00	0.70	0.00
R^2	0.21		0.21		0.22		0.22	
n students	32,504		32,504		50,689		50,689	
n classrooms	4,742		4,742		5,143		5,143	
n schools	437		437		468		468	

Note. Estimates are unstandardized. Outcome variables were standardized, thus estimates are interpretable as change in units of *SD*. *SE* = standard error; *SD* = standard deviation; CELDT = California English Language Development Test; ELP = English language proficiency; FRPL = free or reduced-price lunch eligible; Entered K = student entered U.S. schools in kindergarten or prior; Prop. = proportion; EL = English learner; RFEP = reclassified fluent English proficient; API = Academic Performance Index; ICC = intraclass correlation; R^2 = proportion of variance in the outcome explained by adding fixed factors to a model containing only the intercept and random factors.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 6

The Relation of Classroom Proportion of EL Students and Number of English Language Proficiency Levels With English Learner Student Reclassification to Fluent English Proficient in Elementary School

	RFEP 2012–2013						RFEP 2013–2014					
	Model 1			Model 2			Model 1			Model 2		
	B	*SE*	Exp(*B*)	*B*	*SE*	Exp(*B*)	*B*	*SE*	Exp(*B*)	*B*	*SE*	Exp(*B*)
Intercept	−5.03***	0.55	0.01	−5.03***	0.55	0.01	−3.45***	0.41	0.03	−3.43***	0.42	0.03
Student level												
Prior year CST	1.80***	0.04	6.02	1.80***	0.04	6.03	1.87***	0.04	6.47	1.87***	0.04	6.47
Prior year CELDT	3.47***	0.07	32.06	3.47***	0.07	32.04	3.49***	0.06	32.87	3.49***	0.06	32.86
ELP Level 1	0.32	0.29	1.37	0.31	0.29	1.36	0.37	0.28	1.45	0.37	0.28	1.45
ELP Level 3	0.39**	0.12	1.48	0.40**	0.12	1.49	0.43***	0.10	1.54	0.43***	0.10	1.54
ELP Level 4	0.10	0.13	1.10	0.10	0.13	1.10	0.16	0.12	1.17	0.16	0.12	1.17
ELP Level 5	−1.77***	0.33	0.17	−1.77***	0.33	0.17	−0.11	0.29	0.89	−0.11	0.29	0.90
Female	0.36***	0.05	1.43	0.36***	0.05	1.43	0.36***	0.05	1.43	0.36***	0.05	1.43
Hispanic	0.67	0.41	1.96	0.67	0.41	1.96	−0.01	0.37	0.99	−0.01	0.37	0.99
Asian	−0.38	0.25	0.68	−0.39	0.25	0.68	0.08	0.23	1.08	0.08	0.23	1.08
Spanish home language	−0.92*	0.43	0.40	−0.91*	0.43	0.40	−0.34	0.39	0.71	−0.34	0.39	0.71
FRPL	0.03	0.08	1.03	0.03	0.08	1.03	0.05	0.08	1.05	0.05	0.08	1.05
Entered K	−0.13	0.13	0.88	−0.13	0.13	0.88						
U.S. born	0.06	0.09	1.06	0.06	0.09	1.06	−0.20***	0.09	0.82	−0.20*	0.09	0.82
Grade 4	0.06	0.08	1.06	0.06	0.08	1.06	0.44***	0.08	1.55	0.44***	0.08	1.55
Grade 5	2.42***	0.09	11.24	2.41***	0.09	11.17	2.30	0.09	9.97	2.30***	0.09	9.98
Grade 6	0.09	0.19	1.09	0.09	0.19	1.09	0.36	0.19	1.44	0.37	0.19	1.45

(continued)

Table 6 **(continued)**

	RFEP 2012–2013						RFEP 2013–2014					
	Model 1			Model 2			Model 1			Model 2		
	B	*SE*	Exp(*B*)	*B*	*SE*	Exp(*B*)	*B*	*SE*	Exp(*B*)	*B*	*SE*	Exp(*B*)
Classroom level												
Four to five ELP levels in class	−0.02	0.09	0.98	−0.01	0.09	0.99	−0.08	0.08	0.92	−0.08	0.08	0.92
Prop. EL	0.16	0.17	1.18	0.25	0.19	1.28	−0.03	0.13	0.97	−0.08	0.16	0.93
Four to five ELP levels × prop. EL				−0.41	0.35	0.66				0.12	0.28	1.13
Prop. EL squared				0.21	0.63	1.24				−0.22	0.53	0.80
School level												
Total enrollment	0.00	.00	1.00	0.00	.00	1.00	0.00	.00	1.00	0.00	.00	1.00
Prop. EL	−1.27*	.49	0.28	−1.25*	.50	0.29	0.63	.44	1.88	0.63	.44	1.87
Prop. RFEP	0.45	.68	1.58	0.45	.68	1.56	−0.97	.82	0.38	−0.98	.82	0.38
Prop. FRPL	0.38	.68	1.46	0.37	.68	1.45	1.24	.66	3.45	1.23	.66	3.41
Prop. Hispanic	1.48***	.42	4.40	1.49***	.42	4.45	1.12**	.38	3.06	1.11**	.38	3.04
API	−0.01	.00	0.99	−0.01	.00	0.99	0.00	.00	1.00	0.00	.00	1.00
EL API	0.00	.00	1.00	0.00	.00	1.00	0.00	.00	1.00	0.00	.00	1.00
Title 1	0.56	.50	1.75	0.55	.50	1.74	−0.68	.37	0.50	−0.69	.37	0.50
Similar schools rank	−0.02	.03	0.98	−0.02	.03	0.98						

(*continued*)

Table 6 **(continued)**

	Random Parameters							
	RFEP 2012–2013				RFEP 2013–2014			
	Model 1		Model 2		Model 1		Model 2	
	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*
School level								
Intercept *SD*	0.51	0.05	0.51	0.05	0.53	0.05	0.53	0.00
ICC	0.05		0.05		0.06		0.06	
Classroom level								
Intercept *SD*	1.17	0.05	1.16	0.05	0.95	0.05	0.95	0.05
ICC	0.33		0.33		0.27		0.27	
n students	29,274		29,274		29,049		29,049	
n classrooms	3,369		3,369		2,775		2,775	
n schools	425		425		446		446	

Note. Analyses are cross-sectional. *SE* = standard error; RFEP = reclassified fluent English proficient; CST = California Standards test; ELA = English language arts; CELDT = California English Language Development Test; ELP = English language proficiency; FRPL = free or reduced-price lunch eligible; Entered K = student entered U.S. schools in kindergarten or prior; Prop. = proportion; EL = English learner; RFEP = reclassified fluent English proficient; API = Academic Performance Index; ICC = intraclass correlation.
*$p < .05$. **$p < .01$. ***$p < .001$.

Exploratory Results

Across the 2 years of analysis, we explored the possibility of but found no interaction between classroom proportion EL and number of ELP levels (see Tables 4-6). This finding indicates that the consistently lower average content and language EL performance in classrooms with high proportions of ELs did not vary by number of classroom ELP levels.

Adding a quadratic term for proportion of ELs to the models yielded no statistically significant relationships that replicated across outcomes or across years. However, classroom proportion EL showed a statistically significant negative curvilinear association with 2014–2015 CELDT performance and a similar negative, but not statistically significant, curvilinear association with 2013–2014 CELDT performance. Plotting the relationship showed that after about 0.35 to 0.40 classroom proportion EL, the rate of negative change in EL 2014–2015 CELDT performance accelerated (see Supplemental Figure S1 in the online version of the journal). Plotting the relation between classroom percent EL and 2013–2014 CELDT performance yielded a similar pattern. However, the rate of negative change in CELDT performance accelerated after about 0.15 classroom proportion EL (see the online Supplemental Figure S2).

Research Question 3: What OTL-Related Advantages and Disadvantages Do School Staff Report in Integrated Versus Segregated EL Classroom Compositions. and How Do They Explain Them?

Overview

From the beginning of the study, school staff raised concerns about the rigidity of EL classroom placement policy. In 2011–2012, staff reported that EL placement based solely on ELD level prevented them from considering other student needs and potential that, if permitted, might increase ELs' OTL and promote academic success. A principal explained,

> I would do it more based on what I know the kids need rather than just an ELD level 'cause, I mean, certain kids need certain things from certain—even it might be a certain teacher, but because they're not at ELD 3, I have to put them in this other class . . . because that other teacher has 3s and 4s, but that child is almost there [and] would benefit from everything day in and day out, not just the ELD instructional time, but the core from this other teacher, but I can't do that because it's a different ELD [level]. It would add a third ELD level to that classroom.

Across 2012–2013 and 2013–2014, as District 1 increased the specificity and enforcement of its EL classroom composition policy, staff increasingly and consistently detailed the OTL advantages of integrated classroom compositions in contrast to the disadvantages of segregated compositions. The advantages reported reflected teacher instructional and collaborative

opportunities and student academic and socio-emotional learning opportunities afforded by the heterogeneity of integrated compositions. Conversely, the disadvantages of segregated compositions reflected the parallel loss of opportunities both teachers and students experienced in more homogeneous classrooms segregated by language, which were also segregated by ethnicity and/or race and a wide range of student characteristics, such as performance and behavior. Reports of advantages of segregated compositions were infrequent, and reports of disadvantages of integrated compositions were uncommon. Below we describe staff-reported OTL-related advantages and disadvantages of integrated versus segregated classroom compositions, respectively, which sheds light on the possible mechanisms underlying the quantitative findings.

Theme 1: For Teachers, More Integrated Versus Segregated Classrooms Provided Opportunities for More Effective Instructional Moves

For teachers, it was the presence of English-speaking models and higher performing peers, including EOs, reclassified former ELs, and ELs at higher levels of ELP, in integrated compositions that afforded a greater repertoire of instructional moves, which they considered more effective. Teachers in one school that had traditionally separated ELs and then had integrated them in 2013–2014 for 2 months prior to enforcement of the policy requiring separation had recently experienced both compositions. One teacher explained the difference: "It was such a wonderful thing at the beginning of the year. I had some EOs, RFEPs, IFEPs [initially fluent English proficient students], and the [EL] children—it was just so much—a rich . . . classroom." With this heterogeneity, she could increase small-group instruction:

> I . . . have time to dedicate to a . . . smaller group of children to pull them out to a table—and help them more than having 25 of the same—[because] some of these [other] children will be more independent. . . . I will be more effective if I have a mixture because that way I can reach more children.

As sociocultural and SLA theories would predict, teachers also reported that integrated classroom compositions increased opportunities to structure academic tasks with peer interactions that increased EL academic English language production. A teacher from a different school shared,

> Not just getting interaction from me—in a classroom setting . . . getting the academic language from someone else [English-speaking peers]. . . . I love to partner, think-pair-share, partner talk, group talk, share with your group . . . and then let's share out. I want them to be able to hear individually. I want them to be able to hear others, as well as . . . me model—how to speak and even

> how to listen—
> 'cause that's even a difficult skill—. . . [and] being able to . . . learn from their peers as well.

One teacher who pinpointed the importance of interaction with higher-performing non-EL peers, in addition to the teacher, discussed meeting EL needs in an integrated classroom while increasing academic conversation:

> In terms of the instruction, I'm still aware and keeping those [EL] students in mind—the vocabulary's always emphasized. . . . I'll still provide some sort of sentence frame . . . sentence starter . . . to share [math] problems and being able to discuss the differences in how one person reached one problem [solution] and another person, and maybe seeing and pointing out similarities and differences. So there's always constant conversation—in my math class. When . . . I'm telling them [ELs] to work with a partner . . . and they're not saying much to one another, that's when I have to take a step back and, again, uh, write down the vocabulary . . . word usage and do a lot more scaffolding.

Another teacher contrasted the instructional disadvantages of segregating with the advantages of integrating ELs with higher performing non-EL peers:

> Having all day such a low-ELD group . . . they wouldn't have any examples of good models of writing and reading, and when I ask the higher level thinking questions, they're not able to give me those responses. But they can at least hear . . . that an EO did it, and I'll have them repeat it in a different way . . . so they can practice their vocabulary.

Theme 2: Integrated Classrooms Increased EL OTL Due to Affording a Wider, More Challenging Curricular Focus Requiring Higher Level Thinking and Learning

A teacher expressed the greater possibilities of an integrated composition:

> I would [like] them to be part of my regular class so I can meet their needs. 'Cause it's a whole lot easier to make it an EL-friendly class and not take away the critical thinking and the independent learning as opposed to [in a segregated classroom] bringing them in . . . and they just know when they come to that small table that it's, we're all gonna do the same thing, and it's repeat. The next week it's gonna be all the same thing—[because we are] mandated to use this textbook.

In contrast, segregated EL classrooms narrowed the focus on ELs as students with a language deficit that needed repair. The embedded progress measure was students' ELD standards–based portfolio. A teacher elucidated,

> The teacher's gotta get that ELD portfolio done. And if she's got all 24 [ELs], she'll get it done in increments. It can happen that way, and the teacher will work on, okay, this class is weak in the ELD listening and speaking goals at this [time]. So I'm gonna get those portfolios done. But you can get very myopic, like a horse with blinders, and not see all the rest of it that has to be accomplished for a holistic journey for K through 5.

Theme 3: In Integrated Classrooms, the Diversity/Heterogeneity of Student Peer English Language, Academic, Achievement, and Socio-Emotional Models Increased EL and Non-EL Students' OTL

Staff emphasized the modeling and observational learning opportunities (Bandura, 1986) such peers provided one another. A staff member shared,

> You do have good models for the kids . . . [in integrated compositions]. It just works out so much better because that affective filter is down, and the kids just say whatever they want. They may or may not be corrected, and they don't care, but you get them talking. Again, if you have a classroom where there's a lot of student talking, you're going to get improved language.

Similarly, a teacher focused on peer models and learning from one another,

> My English learners need to associate with their peers and get the language practice from them. Not from adults. Not from books. Not from tapes. . . . They need to converse with their peers. This is how it's naturally been done, and to separate 'em with other English learners, I don't see the advantage. [In integrated classrooms], they [ELs] can practice with a peer who is in that level where they can learn from. They can work in small groups and learn from each other. They can read to a peer that will encourage them to go higher. They will be challenged in this . . . atmosphere. Challenged in a good way, a supportive way, in an environment that will build their confidence.

Ironically, segregated EL compositions produced less rather than more EL English language production, due to teachers' role as the primary English model and talker, according to the teachers. A teacher explained,

> Now that we have all the ELD levels, the lower levels together [again], I think—it's not good for our students. [They] need more modeling than just the teacher . . . especially because . . . we're expecting those children to communicate with each other, and they only have certain amount of [English language], and so we have to be intervening . . . constantly . . . because you're the only English model throughout the day. . . . Sometimes they speak in Spanish to themselves. I say, "Remember we're speaking in English, boys and girls. Let's practice. Let's continue practicing. Now repeat after me. Now repeat." You know, a lot of repetition, but I think if they had other [peer] models,

> they will copy them. . . . "Okay they say it this way. Let me do it." The children feel more comfortable; their affective filter is down . . . the learning will take.

Another teacher said,

> In some cases, there's really nothing they themselves can do, besides what everybody else is doing, you know, what we tell everybody else. "Read more. . . . Reading will improve your writing, and if you speak well"—and, of course, we're always, you know, correcting them and stuff.

Consistent with these views, a nonteaching staff member explicated the negative learning consequences of excessive teacher talk and insufficient student talk in segregated classrooms:

> What I've seen . . . is that you have a teacher doing all the talking. The kids aren't doing any talking, and if you don't speak, you can't read. If you can't read, you can't write. It's a chain reaction.

More broadly, another nonteaching staff member shared her previous experience with segregated EL classrooms:

> Once everybody spoke Spanish, there was no need for it [English]. I noticed that we didn't progress as well in ELD as we should have because we're all communicating, they're learning, but we're not learning English as a second language.

Likewise, for EO student academic and socio-emotional learning, staff asserted, the diversity of integrated classrooms was beneficial, whereas segregation was detrimental. Staff in a school that had gone from integrated to segregated classroom compositions articulated their experience. A nonteaching staff member said,

> So ironically, the students who are in the mainstream, the EOs etc., often take on a particular look, and that's been a big problem for us here at our school. The majority of them [are] African American children, and they just bring different energy. . . . So teachers will say, "We feel like the EO class kind of has shown that it has . . . more verbal students who take away some of the instructional time because of their social issues." So they complain about that. And then our . . . teachers are very upset, 'cause they said it was segregating the class[es]. Like, literally you look in the room, you see all Black kids, and you look next door, and it's all Hispanic children.

A teacher explained,

> It's a little bit difficult . . . this whole like EL and non-EL status . . . at times that [EO-only classroom] comes with some social

> misbehaving—the dynamics of the class, it's just not really healthy. . . . I feel like a class needs to be more diverse . . . you gotta have a little bit of this and a little bit of that—the kids that are EL[s]—they are very anxious to learn, and they're better behaved. If we had mixed it in a little bit more, I think that we could have avoided those problems that we've had this year socially. When the kids are separated like this, I don't think it's fair for those [EO] kids that are in a classroom that's not as socially like together; it's very disruptive to them. . . . I don't know if they're getting the most out of their learning during [this] school year.

Theme 4: Socio-Emotional Benefits Accrued to EL and non-EL Students in Integrated Classrooms, While Staff Worried About the Negative Halo Effect for ELs in Segregated Classrooms

Teachers in the school that had experienced 2 months of integrated classrooms (but had traditionally separated ELs) described the effect on students:

> The children were so happy to help each other. They were talking to each other. My ELD ones were trying to speak more English. . . . I had this RSP [Resource Specialist Program] [EL] student and this little girl who is an IFEP. She was so motivated to see this little girl doing so well in writing. . . . I had 'em, a high with a low. . . . So they could see that it's possible, and I saw that in her writing. . . . When she [EL] came in, she's like, "I can't write a sentence." [The IFEP student said], "Oh yes, you can. You can do it. I know you can. Write me how you can do it." By—just by that, she was motivating this little girl. And you should've seen . . . her [EL student] writing—you could see phonetically, but she knew that she could do it.

Another teacher expounded,

> I had ELD ones . . . newcomers, the EOs, the reclassified, and I just loved . . . it because the EL learners had somebody to model after, not only language . . . academics as well . . . using the—the academic vocabulary. The EL learners hear it. . . . They want to imitate their peers. . . . Of course, that pushes them to do better. Two months into the school year, it's like, "Okay, we're gonna just redo everything [back to segregated EL classroom composition]." . . . Oh my gosh. This was working out so perfect. When the switch started . . . they [ELs] felt like . . . I saw that their demeanor, it's like, "Oh, so and so left." I saw that they were not deflated, but you could see the difference in them. "Oh, like, I got stuck here" kind of thing. . . . Right now, within the groups, they know their level. We just had a few reclassify. [ELs are asking], "Oh, does that mean that he's out [reclassified] of [doing] these portfolio [assignments]? Am I'm almost [out]?" . . . They don't want to be labeled [EL] like that. . . . I don't know if they feel bad. . . . They want to go past that hurdle. They know that it's an achievement.

Similarly, other teachers worried about the negative message sent by segregating ELs from other students. "In English Immersion . . . it's like segregating the youngsters. They don't have a chance to learn from the other kids who are fluent. . . . This is like, 'We are the low-level classes.'"

Theme 5: Classroom-Level EL Segregation Reduced Teacher Collaboration in Both Academic and Socio-Emotional Domains

This practice diminished the availability of teacher peers for meaningful collaboration and had unintended consequences, according to the teachers. When the number of ELs in a grade level resulted in a single segregated classroom, for example, such classrooms were much more likely to have four or five ELP levels, which teachers reported made it difficult to group by ELP level for ELD instruction. In contrast, integrating ELs with non-EL peers in multiple classrooms facilitated teacher collaboration and grouping them for ELD instruction, including regrouping across classrooms as necessary. A teacher explained, "When . . . we actually did have our mixed [integrated] groups [classrooms], but we departmentalized during our ELD instruction block. So one teacher would take on ELDs 1 and 2. The other teacher would take on ELDs 3 and 4." Now with a 100% EL class, she had to provide ELD instruction for Levels 2 to 4 because her peer teacher taught non-ELs only. Similarly, integrated classrooms afforded regrouping by language status for ELD instruction. A teacher who now taught a 100% non-EL class said,

> When I had English learners [in an integrated classroom], . . . we switched. . . . If I have . . . 10 ELs and the other teacher had a certain amount, then . . . during the ELD time, we would [switch]—and then I would instruct those [EL] youngsters, and my [non-EL] kids would, you know, go to the [other] teacher.

Similarly, when students were segregated by language status, diminished teacher peer collaboration often extended to student behavior management as well. A staff member from the school where separating ELs from non-ELs resulted in classrooms with primarily African American or primarily Latino students, explained the detrimental effects on teacher collaboration, "Behavior tends to be clustered when we separate the classes based on language need. There isn't a balance or the flexibility for teachers to collaborate and share the social or behavior concerns."

Theme 6: Diminished Instructional and Collaborative Opportunities as Well as Extra Work in Segregated EL Classrooms Took a Toll on Teachers

A nonteaching staff member said,

> Teachers that have the BCLAD [and are assigned to teach EL students] many times get very frustrated because they're . . . maybe in fourth grade [and have] all the lower ELD-level [ELs]. And . . . if the children are staying at low ELD levels, it's reflected in other subject areas. They want the opportunity to teach a class with different dynamics, per se.

A teacher indicated,

> It's a challenge to just have the lower EL levels. I love teaching them, but then they [administration] expect them to achieve. . . . You're not checking [that] these [students] are ELD 3. I mean, who looks at that? You just look at scores. That's when you feel like, "I don't wanna teach this." If you're just comparing numbers, it's, it's not fair.

Another teacher who had taught a segregated EL class of mostly ELP Level 1 students described becoming demoralized:

> It was a small class. I must have had about 20 kids, and they did that to see if these kids can raise their levels, and it was not that they didn't want to. It's they couldn't. I saw very little progress in that class. It beats you down because you don't see success.

Similarly, another teacher described the isolation of students and teachers in segregated EL classrooms:

> There's a frustration in the teachers that end up with the wall-to-wall [EL] class . . . because they need models. . . .There's nobody built in . . . to help you. . . . So their only model is the teacher. . . . So they don't get the social interaction [either]. They don't, immerse. They're apart . . . then you're working against that [too].

Theme 7: Advantages of Segregated EL Classroom Compositions and Disadvantages of Integrated Compositions Were Scarce

Advantages of segregated classrooms were focused almost exclusively on easier planning, and no staff members reported that it translated into more effective instruction. One teacher explained, "Like the planning and the prepping . . . if it's homogeneous, in a sense, it is a little easier, 'cause it's less differentiated planning that I may have to do." Another said, "I think it would just be easier to plan." Similarly uncommon, staff-reported disadvantages of integrated compositions focused on the possibility that non-EL peers might dominate or ELs might get overlooked when teachers were unprepared to meet their needs. A nonteaching staff said, "In a classroom that was heterogeneous, if the teacher did not have enough strategies or skill to work with EL students, they could get ignored."

Research Question 4: How Do the Quantitative and Qualitative Findings Corroborate or Help Explain Each Other?

The consistent quantitative findings that ELs exhibit lower language and content performance in more segregated classrooms corroborated school staff reports that overwhelmingly described OTL advantages in integrated classroom compositions versus disadvantages in segregated classroom compositions. Analogously, staff reports corroborated the quantitative findings and, additionally, shed light on the possible mechanisms underlying the quantitative patterns. Across methods, the findings also cohered with our theoretical framework, indicating that who is in the classroom matters because it defines the pool of eligible teachers and peers for instruction, learning, language, and relationships. Staff emphasized the teacher instructional opportunities and student academic and socio-emotional opportunities to learn afforded by the language, performance, behavioral, and ethnic/racial diversity of integrated classrooms. Conversely, staff described how segregated classrooms diminished these opportunities and how they reduced teacher collaboration, increased EL teacher burnout, and fostered low EL status and stigma. Below we discuss these points in more detail.

Discussion

District 1's well-intentioned classroom composition policy—which directed schools to create separate classrooms for ELs at ELP Levels 1–3 while maintaining no more than three adjacent ELP levels in classrooms and to backfill with ELs at ELP Levels 4 and 5 if necessary to achieve 100% EL classrooms—was premised on the rationale that increasing homogeneity would raise teachers' capacity to target EL needs, thus improving instructional effectiveness and achievement. Neither the quantitative nor the qualitative evidence, which were mutually corroborative, supported this view, however. Rather, the quantitative evidence revealed that more highly segregated EL classroom compositions were associated with depressed EL performance on state tests of ELP, ELA, and math standards, consistent with our hypothesis and the sparse previous EL-specific research (e.g., Abedi & Herman, 2010). We found suggestive evidence that the negative relationship with ELP performance accelerates with increasing EL classroom concentration; this awaits further research. Notably, the negative relationship between classroom-level EL segregation and performance did not vary by classroom number of ELP levels. Similarly, we found almost no evidence that maintaining up to three adjacent ELP levels in classrooms versus four to five was advantageous, thus our exploratory hypothesis was largely unsupported. Neither classroom proportion EL nor number of ELP levels was associated with reclassification. These latter results were not

surprising given that meeting reclassification criteria in the current year largely reflected students' previous year's performances on state tests.

Corroborating these quantitative findings, staff-reported qualitative evidence illuminated potential underlying classroom-level mechanisms consonant with sociocultural and SLA theories. They highlighted teacher instructional opportunities and student OTL benefits in integrated (vs. segregated) classrooms composed of students with diverse/heterogeneous language, performance, socio-emotional, and ethnic/racial backgrounds. For teachers, integrated classrooms afforded a wide range of instructional moves that capitalized on having English language, academic language, and academic work models. These included increasing EL English language production and academic work via a variety of organizational structures such as peer and small-group work, discussion, and sharing, which are among the NASEM (2017)–recommended practices for EL language and content learning. Teachers also zeroed in on EL opportunities for engaging in increased and more advanced levels of academic discourse and learning with higher performing, English-proficient peers in integrated classrooms—bolstering the notion that meaningful discourse within the social plane is the cornerstone of language and content learning (e.g., Tharp et al., 2000). Integrated classrooms with a diverse set of students, some of whom could work independently, also increased time for teacher-led, small-group targeted instruction. Finally, echoing previous qualitative reports (Lillie et al., 2010; Rios-Aguilar et al., 2012), teachers underscored that integrated classrooms fostered more challenging curricular foci.

For EL students, staff emphasized the language, academic, and socio-emotional learning opportunities that heterogeneous peer models provided in integrated classrooms. For EL language and content learning, staff underscored the necessity of engaging in talk and academic tasks with higher-performing, English-speaking peers, not just teachers. Such contexts lowered ELs' affective filter for language production, increased student versus teacher talk, fostered mutual peer learning, and paved the way for improving reading and writing. In contrast, segregated EL classrooms resulted in less EL English language production due to teachers functioning as the primary English language and/or academic English model. Likewise, for non-EL students, staff stressed the academic and socio-emotional learning benefits of EL peer models in integrated classrooms. Clustering of social and behavioral issues within segregated classrooms also diminished student OTL and learning. For ELs and non-ELs alike, then, in contrast to integrated classrooms, the constraints of segregated classrooms meant less exposure to and opportunity to learn from a range of language, performance, and socio-emotional models.

Finally, the evidence pointed to two other unintentional consequences of student segregation. First, as predicted by labeling theory, the deficit EL label, combined with policy requiring separation and highlighting the need for remedy, coincided with the conditions and processes that foster low status and

stigmatization, along with a cascade of constraints for school staff and ELs alike. Classroom-level language status separation translated into segregation by ethnicity and race and a wide range of student characteristics, such as performance and behavior, according to school staff. For principals, implementing the policy separating ELs from non-ELs meant that language status and, for ELs, ELP level drove classroom placement. Occluded from view and consideration was ELs' full range of academic and socio-emotional needs and potential. Teachers too indicated that segregated EL classrooms narrowed their focus on ELs as students with a language deficit, restricted their instructional repertoires, and reduced EL OTL. Expressing worry about EL awareness of what the label's low status combined with segregation signaled, teachers described ELs' desire to shed their embodiment of the label and placement in the "low class." Second, without a shared range of language, performance, and socio-emotional strengths and needs across classrooms, teacher collaboration also diminished, while workload, demoralization, and burnout increased among teachers of segregated EL classrooms. Both students and teachers appeared to suffer under these circumstances. Thus, the findings cohered with theory and across methods: Quantitative findings revealed broad patterns, while qualitative findings pointed to possible underlying mechanisms.

Research and Policy Implications

The findings underscore that classroom composition is a powerful policy lever that districts must often manipulate in the context of scarce empirical evidence. Urban students often reside in racially/ethnically, linguistically, and socio-economically segregated neighborhoods. Our findings, which cohered with theory and across methods, indicate that further segregating ELs in classrooms within elementary schools may constrain EL OTL, is negatively associated with EL achievement, and may also foster low status and stigma. Without evidence to the contrary, such action appears unwarranted. Our qualitative findings suggest that segregated classrooms can also be detrimental for non-EL peers.

We also uncovered evidence of parallel, diminished teacher opportunities for more effective instruction and peer collaboration, and increased burnout in segregated classrooms. Notably, the hallmarks of successful EL programs include major shifts in adult mind-sets from needing to remediate ELs' lack of English proficiency to focusing on academic development based on diagnosing needs and addressing them instructionally; high expectations for students; and rigorous, grade-level curriculum and instruction rather than those geared to ELP level (NASEM, 2017). Yet the conditions and constraints school staff and ELs experienced under classroom-level segregation policies highlighting the need for remedy appeared to foster adult mind-sets, expectations, and practices opposite those characteristics. What appears to matter both to students and to teachers are the sociocultural resources they experience and the affordances these resources provide. In combination, these

findings suggest that classroom composition policies are implicated in the overall ecology of schools.

Integrating our quantitative and qualitative results, we hypothesize that classroom OTL factors, such as ELs' access to English-proficient peers and teachers' increased opportunities for more effective instructional moves, may mediate the relation between classroom proportion EL and EL outcomes. In combination, the findings suggest that, on average, placing ELs in classrooms with higher performing, English-proficient non-EL peers will more likely facilitate their achievement and serve them best. In addition to being statistically significant, our estimates of 1/6 to 1/5 standard deviation lower average EL achievement in more segregated versus integrated classrooms are practically meaningful. For ELs, even 1 scale score point can mean the difference between reclassifying or not, and the often concomitant access to increased OTL (e.g., see Callahan & Shifrer, 2012; Estrada, 2014, 2018). In our data, these differences in performance translated to 6 to 9 points on the CST-ELA, depending on grade level, and about 6 points on the CELDT for every grade level. Our findings cohere with Abedi and Herman's (2010) findings that Grade 8 math classrooms with higher proportions of ELs were associated with lower levels of both OTL and performance. They are also congruous with research showing that segregating ELs produced isolation and less rigorous instruction for ELs (Gándara & Orfield, 2012; Lillie et al., 2010; Ríos-Aguilar et al., 2012). Moreover, they are consistent with the general literature indicating that placing elementary students in homogeneous classrooms by performance levels does not lead to improved performance (e.g., Slavin, 1989). Additionally, they support Saunders and Goldenberg's (2010) recommendation against classroom-level EL segregation. Our findings begin to fill the gaps in EL-specific research and the evidence gap for developing classroom composition policy that supports EL school success. Taken together, the weight of the available evidence indicates that maximizing integration is likely to support increased EL achievement, on average.

Strengths of our study include using mixed methods across multiple years and grade levels to discern broad patterns with quantitative data, identifying possible underlying mechanisms with qualitative data, examining the extent to which the findings cohere across methods and with theory, and integrating multiple theories to deepen understanding of interrelationships and generate hypotheses. To our knowledge, this is the first study to do so in this area of research. Simultaneously, our study has several limitations. Despite accounting for student-, classroom-, and school-level covariates likely to influence both classroom placement and outcomes in the quantitative analyses, there might be additional, unmeasured variables that might explain the relationships between classroom composition and outcomes; therefore, we cannot infer causality. Second, the quantitative findings pertain to EL student achievement only. Yet the qualitative findings not only corroborate that for ELs segregated versus integrated classroom compositions are

disadvantageous, they likewise suggest that segregated classrooms can be detrimental for non-ELs too. Third, we relied on self-reports of instruction and OTL in a small sample of schools. Fourth, although we were restricted to a single district within a single state, we were able to capitalize on the natural variation in classroom composition in one of the largest districts in the country with a very large EL population. Finally, our results pertain primarily to English-only instructional contexts; yet ELs nationwide most commonly experience English-only instruction. We are mindful that dual-immersion programs within highly segregated schools can succeed with ELs (e.g., Lindholm-Leary & Block, 2010). We suspect that in these settings positive results are likely due to multiple factors: (a) language and literacy development in the first language (L1) prior to or concurrent with L2 and (b) retaining higher performing students who meet ELP and content performance criteria in these classrooms, thus mirroring more integrated classroom compositions, which our findings indicate afford more OTL.

At a time when, reflecting society, our schools are becoming increasingly diverse, it is crucial that there is further mixed-methods research employing causal and quasi-experimental research designs implemented across a variety of contexts, using direct measures of instruction, OTL, and performance outcomes for all students, along with concomitant qualitative data such as staff interviews. Such an approach will produce more detailed evidence regarding broad patterns and underlying mechanisms, which will deepen our understanding and ideally advance our interpretations in the direction of causality. Finally, such research holds promise for developing evidence-based policy for nurturing the achievement and well-being of ELs, the fastest-growing segment of K–12 students nationwide.

Notes

We thank Claude Goldenberg for thoughtful feedback on this article, Aleshia Barajas and Soyoung Park for research assistance, and our district partners and the school staff who made our work possible and enriched our understandings.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education through Grant R305A11512 to The Regents of the University of California, Santa Cruz. The opinions expressed are those of the authors and do not represent the views of the Institute of Education Sciences or the U.S. Department of Education. The memorandum of understanding with the participating district does not permit us to share the data or samples.

[1]California's English-only instructional programs include (a) Structured English Immersion—for ELs not yet reasonably fluent as defined by the district—in which nearly all instruction is in English and curriculum and presentation are for children learning English, and (b) English Language Mainstream—for ELs who are reasonably fluent. In both programs, until they reclassify as fluent English proficient (FEP), ELs continue to receive English language development (ELD) instruction and additional educational services to recoup any academic deficits due to language barriers (Sugarman, 2018).

[2]The zone of proximal development is the distance between a performance capacity made possible by the assistance of a more expert other and unassisted performance, an independent performance capacity (Vygotsky, 1978).

[3]Cognitivist SLA perspectives emphasize information processing in language learning (Atkinson, 2011). Language socialization SLA perspectives emphasize that language learning occurs through interaction with more knowledgeable interlocuters, with a focus on the micro- and macrocontexts in which language learning and use occur and on language socialization (Duff & Talmy, 2011).

References

Abedi, J., & Herman, J. (2010). Assessing English language learners' opportunity to learn mathematics: Issues and limitations. *Teachers College Record*, *112*, 723–746.

Atkinson, D. (2011). Introduction: Cognitivism and second language acquisition. In D. Atkinson (Ed.), *Alternative approaches to second language acquisition* (pp. 13–35). New York, NY: Routledge.

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

Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs, NJ: Prentice Hall.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77–101.

Burton, L., & Welsh, W. (2015). *Inequality and opportunity: The role of exclusion, social capital, and generic social processes in upward mobility*. New York, NY: William T. Grant Foundation. Retrieved from https://wtgrantfoundation.org/library/uploads/2016/01/Inequality-and-Opportunity-Burton-and-Welsh-William-T.-Grant-Foundation.pdf

Calhoon, M. B., Al Otaiba, S., Cihak, D., King, A., & Avalos, A. (2007). Effects of a peer-mediated program on reading skill acquisition for two-way bilingual first-grade classrooms. *Learning Disability Quarterly*, *30*, 169–184.

California Department of Education. (2012). *Dataquest*. Retrieved from http://data1.cde.ca.gov/dataquest/

California Department of Education. (2015). California English Language Development Test: Technical Report 2014-15 Edition. Retrieved from https://www.cde.ca.gov/re/pr/documents/celdt1415techrpt.pdf

California Department of Education. (2019). *The Public Schools Accountability Act of 1999: CalEdFacts*. Retrieved from https://www.cde.ca.gov/ta/ac/pa/cefpsaa.asp

Callahan, R. M., & Schifrer, D. R. (2012). High school ESL placement: Practice, policy, and effects on achievement. In Y. Kanno & H. Harklau (Eds.), *Linguistic minority students go to college* (pp. 19-37). New York, NY: Routledge.

Carroll, J. B. (1972). Defining language comprehension: Some speculations. In R. O. Freedle & J. B. Carroll (Eds.), *Language comprehension and the acquisition of knowledge* (pp. 1–29). New York, NY: Wiley.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Thousand Oaks, CA: Sage.

Dabach, D. B. (2014). "I am not a shelter!": Stigma and social boundaries in teachers' accounts of students' experience in separate "sheltered" English learner classrooms. *Journal of Education for Students Placed at Risk*, *19*, 98–124. doi:10.1080/10824669.2014.954044

DeKeyser, R. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 42–63). Cambridge, England: Cambridge University Press.

Duff, P. A., & Talmy, S. (2011). Language socialization approaches to second language acquisition: Social, cultural, and linguistic development in additional languages. In D. Atkinson (Ed.), *Alternative approaches to second language acquisition* (pp. 107–128). New York, NY: Routledge.

EdSource. (2008). *English learners in California: What the numbers say*. Mountain View, CA: Author. Retrieved from https://files.eric.ed.gov/fulltext/ED501201.pdf

Ellis, R. (2005). Principles of instructed language learning. *System*, *33*, 209–224.

Epstein, J. L. (1983). The influence of friends on achievement and affective outcomes. In J. L. Epstein & N. Karweit (Eds.), *Friends in school: Patterns of selection and influence in secondary schools* (pp. 177–200). New York: Academic Press.

Estrada, P. (2014). English learner Curricular Streams in four middle schools: Triage in the trenches. *Urban Review*, *46*, 535–573. doi:10.1007/s11256-014-0276-7

Estrada, P. (2018). Secondary English learner Curricular Streams, opportunity to learn, and academic outcomes. In D. Conger (Chair), *English learner course taking and academic success*. Symposium conducted at the annual meeting of the American Association of Policy Analysis and Management, Washington, DC.

Estrada, P., Park, S., & Farkas, T. (2018, April). *Conceptions of secondary English learners: "They call them lifers."* Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Estrada, P., & Wang, H. (2018). Making English learner reclassification to fluent English proficient attainable or elusive: When meeting criteria is and is not enough. *American Educational Research Journal*, *55*, 207–242. doi:10.3102/0002831217733543

Fillmore, L. W., & Snow, C. (2000). What teachers need to know about language. In C. T. Adger, C. E. Snow, & D. Christian (Eds.), *What teachers need to know about language* (pp. 7–53). McHenry, IL: Delta Systems/Center for Applied Linguistics.

Flores, S. M., Batalova, J., & Fix, M. (2012). *The educational trajectories of English learners in Texas*. Washington, DC: Migration Policy Institute.

Gándara, P., & Orfield, G. (2012). Segregating Arizona's English learners: A return to the "Mexican room"? *Teachers College Record*, *114*(9), 1–27.

Gándara, P., Rumberger, R., Maxwell-Jolly, J., & Callahan, R. (2003). English learners in California schools: Unequal resources, unequal outcomes. *Education Policy Analysis Archives*, *11*(36), 1–52.

Gersten, R., Baker, S. K., Shanahan, T., Linan-Thompson, S., Collins, P., & Scarcella, R. (2007). Effective literacy and English language instruction for English learners in the elementary grades (*IES Practice Guide*, NCEE 2007-4011). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee/pdf/20074011.pdf

Goldenberg, C. (2013). Unlocking the research on English learners: What we know—and don't yet know—about effective instruction. *American Educator*, *37*, 4–11.

Gottfried, M. A. (2014). Classmates with disabilities and students' noncognitive outcomes. *Educational Evaluation and Policy Analysis*, *36*, 20–43.

Greenfader, C. M., Brouillette, L., & Farkas, G. (2015). Effect of a performing arts program on the oral language skills of young English learners. *Reading Research Quarterly*, *50*, 185–203.

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, *18*, 527–544.

Harklau, L. (2000). From the "good kids" to the "worst": Representations of English language learners across educational settings. *TESOL Quarterly*, *34*(1), 35–67. doi:10.2307/3588096

Hoxby, C. (2000). *Peer effects in the classroom: Learning from gender and race variation* (NBER Working Paper No. w7867). Cambridge, MA: National Bureau of Economic Research.

Lillie, K. E., Markos, A., Estrella, A., Nguyen, T., Trifiro, A., Arias, M. B., . . . Pérez, K. (2010, July 8). Policy in practice: The implementation of Structured English Immersion in Arizona. *Civil Rights Project/Proyecto Derechos Civiles*. Retrieved from https://www.civilrightsproject.ucla.edu/research/k-12-education/language-minority-students/policy-in-practice-the-implementation-of-structured-english-immersion-in-arizona

Lindholm-Leary, K., & Block, N. (2010). Achievement in predominantly low SES/Hispanic dual language schools. *International Journal of Bilingual Education and Bilingualism*, *13*, 43–60.

Link, B. G., & Phelan, J. C. (2001). Conceptualizing stigma. *Annual Review of Sociology*, *27*, 363–385.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition: Vol. 2. Second language acquisition* (pp. 413–468). New York, NY: Academic Press.

McDonnell, L. M. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis*, *17*, 305–322. doi:10.3102/01623737017003305

National Academies of Sciences, Engineering, and Medicine. (2017). *Promoting the educational success of children and youth learning English: Promising futures*. Washington, DC: National Academies Press. doi:10.17226/24677

O'Day, J. (2009). Good instruction is good for everyone—or is it? English language learners in a balanced literacy approach. *Journal of Education for Students Placed at Risk*, *14*, 97–119.

Pennycook, A. (2002). *English and the discourses of colonialism*. New York, NY: Routledge.

Pennycook, A. (2007). *Global Englishes and transcultural flows*. London, England: Routledge.

Plano Clark, V. L., & Badiee, M. (2010). Research questions in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social and behavioral research* (pp. 275–304). Thousand Oaks, CA: Sage.

Ríos-Aguilar, C., González-Canche, M. S. G., & Moll, L. C. (2012). Implementing Structured English Immersion in Arizona. *Teachers College Record*, *114*(9), 1–18.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. Oxford, England: Oxford University Press.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67.

San Diego Unified School District. (2005). *Administrative procedure: Instructional programs for English learners*. Retrieved from https://www.sandiegounified.org/sites/default/files_link/district/files/procedures/pp4250.pdf

San Jose Unified School District. (2018). *Education for English language learners* (BP 6174). Retrieved from http://konoctiusd.org/wp-content/uploads/2017/12/6174-BP-AR-E-Education-for-English-Language-Learners.pdf

Saunders, W. M., & Goldenberg, C. (2010). Research to guide English language development instruction. In *Improving education for English learners: Research-based approaches* (pp. 21–81). Sacramento, CA: California Department of Education.

Saunders, W. M., Goldenberg, C., & Marcelletti, D. (2013). English language development: Guidelines for instruction. *American Educator*, *34*(2), 13–25.

Saunders, W. M., & Marcelletti, D. J. (2013). The gap that can't go away: The catch-22 of reclassification in monitoring the progress of English learners. *Educational Evaluation and Policy Analysis*, *35*, 139–156.

Schisterman, E. F., Cole, S. R., & Platt, R. W. (2009). Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology*, *20*, 488–495.

Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research*, *57*, 293–336.

Slavin, R. E. (1989). Ability grouping and student achievement in elementary schools. In R. E. Slavin (Ed.), *School and classroom organization* (pp. 129–151). Hillsdale, NJ: Lawrence Erlbaum.

Sugarman, J. (2018, June). *A matter of design: English learner program models in K–12 education* (Issue Brief No. 2). Washington, DC: Migration Policy Institute.

Tharp, R., Estrada, P., Dalton, S., & Yamauchi, L. (2000). *Teaching transformed: Achieving excellence, fairness, inclusion, and harmony*. Boulder, CO: Westview.

Tharp, R. G., & Gallimore, R. (1988). *Rousing minds to life*. New York, NY: Cambridge University Press.

Thompson, K. D. (2015). Questioning the long-term English learner label: How classification and categorization can blind us to students' abilities. *Teachers College Record*, *117*, 1–50.

Umansky, I. M. (2016). To be or not be EL: An examination of the impact of classifying students as English learners. *Educational Evaluation and Policy Analysis*, *38*, 714–737.

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children*, *23*, 34–41.

Wolf, M. K., Kao, J., Griffin, N., Herman, J. L., Bachman, P. L., Chang, S. M., & Farnsworth, T. (2008). *Issues in assessing English language learners: English language proficiency measures and accommodation uses. Practice review (Part 2 of 3)* (No. 732). Los Angeles, CA: CRESST.

Zhang, J., Anderson, R. C., & Nguyen-Jahiel, K. (2013). Language-rich discussions for English language learners. *International Journal of Educational Research*, *58*, 44–60.

Manuscript received January 25, 2019
Final revision received August 6, 2019
Accepted October 2, 2019