Executive functions in two-way dual-language education: A mechanism for academic performance

Alena G. Esposito

BILINGUAL RESEARCH JOURNAL
https://doi.org/10.1080/15235882.2021.1874570

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

Executive functions in two-way dual-language education: A mechanism for academic performance

Alena G. Esposito

Clark University, USA

Children across the United States are increasingly learning academic content through two-way dual-language education (http://www.cal.org/twi/). This education model provides instruction through two languages in classrooms comprised of approximately equal numbers of native and non-native English speakers. For both language groups, this educational model is an effective approach for achieving second-language fluency (García & Náñez, 2011; Lindholm-Leary & Genesee, 2014). Importantly, both native and non-native English speakers in dual-language education programs perform as well or better academically than their peers in mainstream English classrooms (e.g., Marian, Shook, & Schroeder, 2013; Steele et al., 2017). However, the mechanisms that explain this academic advantage remain to be understood. We examined the possibility that enhanced executive functions through second-language exposure underlie the academic benefits of dual-language education in a rural, low-income, sample of elementary school students.

Dual-language education and participating school system

Bilingual education is an umbrella term that encompasses two-way dual-language education, one-way dual-language education, and immersion education models among others. Dual-language education models teach academic content through two languages, such that children learn both the languages and the content as they progress through school. The 3 primary goals for dual-language programs are to support academic achievement, develop bilingualism and biliteracy, and foster socio-cultural competence (Howard et al., 2018). In two-way dual-language education, classes aim for a 50/50 composition between speakers of the paired languages (language pairings vary; English/Spanish is common in the U.S.) and at least 50% of content provided through the partner language (some models provide up to 90%, initially, tapering down to 50%). Two-way dual-language differs from one-way dual-language in the 50/50 composition of speakers. Dual-language differs from immersion models in that there is typically less content provided through the partner language (often 90–100% and typically comprised non-native speakers of that language; e.g., English speakers in a French immersion in Montreal). Thus, dual-language refers to both one-way and two-way programs and bilingual education is a broader term encompassing dual-language and immersion models.

The specific collaborating school system for the reported work developed a two-way dual-language program as an optional strand within the schools to address the needs of their community (approximately one-third[rd] are native Spanish speaking). In this particular model, instruction alternates days between Spanish and English, thus providing a 50–50 model of content instruction through each language. Children enter through two lotteries, one for native English speakers and one for native Spanish speakers, ensuring the 50/50 composition of native speakers of each of the partner languages.

CONTACT Alena G. Esposito Aesposito@clarku.edu Clark University, 950 Main Street, Worcester, MA 01610, USA.
Dr. Alena G. Esposito is a developmental psychologist focused on cognitive development. Her research concentration is malleable factors influencing learning and subsequent academic achievement. Esposito is a former elementary educator who earned her Ph.D. in Lifespan Developmental Psychology from North Carolina State University.

Students not in the program are placed in a mainstream English education model within the same school.

Dual-language education and academic performance

Multiple studies indicate that children in bilingual education models (including dual-language and immersion models) have academic outcomes that match or even exceed those of their peers in mainstream education models, especially in later elementary grades, (e.g., Cobb, Vega, & Kronauge, 2006; Lindholm-Leary & Genesee, 2014; Padilla, Fan, Xu, & Silva, 2013; Steele et al., 2017). For example, Marian et al. (2013) investigated the academic achievement of students in grades 3, 4, or 5 (approximately ages 8–10 years), a portion of which were enrolled in a two-way dual-language program. They found an advantage in academic performance in math across all three grade levels and reading in 3rd grade. Similarly, Watzinger-Tharp, Swenson, and Mayne (2018) examined growth in over 2000 4th grade students in either mainstream English education or a dual-language education model (comprised of both one-way and two-way models across three partner languages). In a matched-sample of mainstream and dual-language students, the dual-language students showed greater growth in math achievement across the 4th grade year.

Despite seemingly robust evidence for an academic advantage for bilingual education participants, the effect is still in question. In a meta-analyses of 10 studies reporting academic performance for students in one-way or two-way dual-language programs compared to mainstream programs, Hill (2018) determined the effect to be null. The results indicated a small positive effect that Hill proposed could be easily nullified by the inclusion of a few studies with even small negative effects. He also questioned whether the reported results that show an advantage to those in dual-language education are due to their participation or can be explained by changing demographics as a result of attrition. Attrition, he reports, is likely to positively affect the socio-economic status of the group because low-income families are more transitory and more likely to relocate, leaving predominantly higher socio-economic status students in the program. Socio-economic status is a known predictor of academic performance (e.g., Hoff, 2013; Nesbitt, Baker-Ward, & Willoughby, 2013). The question of attrition also arises as students who are struggling academically may be more likely to leave the program for mainstream education models. It is, therefore, important to examine academic achievement with consideration for student intelligence and family socio-economic status.

Dual-language education and executive functions

The "bilingual advantage" refers to higher performance by bilingual compared to monolingual individuals in executive functions (for a review, see Bialystok, Craik, & Luk, 2012). Executive functions are the top-down processes that are required for effortful cognition such as reasoning, problem-solving, and planning and include the core components of inhibition, interference control, working memory, and cognitive flexibility (Diamond, 2013). Executive functions are positively correlated to both socio-economic status (SES; e.g., Nesbitt et al., 2013) and academic performance (for review, see Serpell & Esposito, 2016). Bilingualism may be a protective factor for children from low-SES backgrounds, providing an advantage that offsets the disadvantages associated with their economic conditions. The bilingual advantage is thought to result from constant practice in managing two languages, which enhances mental flexibility and controlled attention. The advantage has been found across age groups, languages, and geographic locations. However, the specific conditions under which a bilingual advantage is and is not found have not been elucidated (e.g., Yang, Hartanto, & Yang, 2016; Valian, 2015). While some research shows benefits to executive functions after short periods of intense training (e.g., Janus, Lee, Moreno, & Bialystok, 2016), others show that benefits only emerge after a threshold of proficiency in both languages is met (e.g., De Cat, Gusnanto, & Serratrice, 2018). Thus, the second-language exposure gained through the specific context of two-way dual-language education may not be sufficient for an executive functions advantage to develop.

The research investigating whether an advantage develops for children in bilingual education models has found mixed results. Studies examining immersion education have found support for emerging benefits to executive functions after 3 or more years of participation (e.g., Bialystok & Barac, 2012; Nicolay & Poncelet, 2013, 2015). In contrast, several studies have failed to find differences between two-way dual-language and mainstream monolingual education models (e.g., Kaushanskaya, Gross, & Buac, 2014; Poarch & van Hell, 2012) with at least one study finding a disadvantage (Purić, Vuksanović, & Chondrogianni, 2017). These studies, however, examined children with less than 2 years of experience in dual-language programs and, where reported, children were from middle-class backgrounds. Hartanto, Toh, and Yang (2018) found that the bilingual advantage was only evident for children from low-SES backgrounds. Thus, if the conditions of two-way dual-language education are in-line with developing a bilingual advantage, it is more likely to emerge for children who have more than 2 years of experience in the program and are from low-SES backgrounds who can most benefit from an intervention. In support of this hypothesis, a study examining controlled attention for children in an area of marked poverty enrolled in either mainstream monolingual education or two-way dual-language education for more than 3 years found evidence for emerging benefits (Esposito & Baker-Ward, 2013).

As documented in an extensive literature, executive functions correlate with indices of academic performance (for review, see Serpell & Esposito, 2016). For example, Best, Miller, and Naglieri (2011) measured academic achievement and executive functions in a nationally representative sample of children aged 5–17 years and found a consistent relation between executive functions performance and academic achievement in math and reading. Thus, if two-way dual-language education conveys a benefit to executive functions, that benefit could support greater academic achievement.

The present study

We examined whether enhanced executive functions gained through participation in a two-way dual-language program are a mechanism through which an academic advantage emerges. In sum, there is evidence for an academic advantage in bilingual education models, but the advantage is still in question and the mechanism is unknown. We propose that two-way dual-language education fosters executive functions similar to the advantage found in bilingual individuals and that well-developed executive functions are a mechanism for an academic advantage. In the present cross-sectional study, we recruited a sample of primary and intermediate elementary students in either two-way dual-language education (Spanish/English) or mainstream English education in an area of rural poverty within the same schools. The participating school system provided academic data. Parents provided information about the home environment and family demographics. We met with students individually to measure executive functions as well as variables that permitted us to create a matched-sample of students in two-way dual-language education and the mainstream English model. The full-sample provided a larger sample size and included all participants, including those on the ends of the score distribution. The matched-sample, however, allowed us to address three criticisms of previous literature, namely that few studies equated groups on individual intelligence, family socio-economic status, and sample size.

The present study had four research questions: 1) is there an academic advantage for children enrolled in two-way dual-language compared to mainstream English education in an area of rural poverty; 2) is the second-language experience provided through a 50/50 two-way dual-language education model sufficient to benefit executive functions; and, if so, 3) is there evidence that executive functions are a mechanism through which the academic advantage for two-way dual-language participants emerges? Relatedly, 4) does the pattern of results in a full-sample analyses replicate in a matched-sample controlling for participant intelligence and family demographics? We predicted that we would replicate the dual-language academic advantage in late elementary students, especially in math where the findings of an advantage appear to be the most robust. We also predicted a dual-language advantage in executive functions for late elementary students. We predicted that executive functions would mediate the relation between educational model and academic performance, providing evidence for a mechanism of the academic advantages found in models of bilingual education. We predicted the reduced sample size and elimination of the extreme ends of the score distributions of the matched-sample would reduce the power and effect sizes of the results, but that the pattern of results would be similar to those found in the full-sample.

Method

Participants and school system

The participants were 288 children (primary = grades K-1, *n*= 175, Mean age = 6 years, 11 months; intermediate = grades 4–5, *n*= 113, Mean age =10 years, 9 months), enrolled in a rural public school system in the southeastern United States (see Table 1). The matched-sample was a subset exact matched on grade and we used Coarsened Exact Matching (CEM) on verbal and non-verbal intelligence as well as parent/guardian education level (*n*=136). The school system includes 266 mi^2 and all students within the county attended the same 2 schools as part of a continuous progression through the county grade school program (K-2 primary school, early elementary; and 3–5 intermediate school;

Table 1. Full Sample and Matched Sample by School.

	Primary				Intermediate			
	Dual-Language		Mainstream		Dual-Language		Mainstream	
	Full	CEM	Full	CEM	Full	CEM	Full	CEM
n (female)	50 (27)	42 (22)	125 (62)	42 (16)	34 (22)	26 (13)	79 (36)	26 (17)
Mean WASI Vocabulary (*SD*)	19.56 (9.99)	19.13 (9.43)	18.82 (8.94)	19.23 (9.17)	38.00 (8.16)	37.62 (7.74)	35.38 (6.58)	36.50 (6.74)
Mean WASI Blocks (*SD*)	8.67 (6.08)	7.10 (4.74)	6.59 (4.58)	7.28 (4.92)	21.48 (12.01)	21.29 (12.20)	16.95 (10.14)	18.65 (10.19)
Caregiver Education Level (*SD*)	4.33 (1.78)	4.38 (1.86)	4.08 (1.49)	4.05 (1.55)	4.43 (2.02)	4.61 (1.94)	4.26 (1.62)	4.29 (1.27)
Hours Spent with an Adult on Homework	3.32 (1.78)	3.31 (1.83)	3.38 (2.31)	3.40 (2.41)	2.13 (1.78)	2.37 (1.86)	2.95 (2.25)	2.64 (1.94)
Hours Spent with an Adult Reading	2.94 (1.81)	3.08 (1.90)	3.20 (2.17)	3.46 (2.49)	1.52 (1.17)	1.47 (1.17)	2.45 (2.05)	2.44 (2.06)
Participation in extra-curriculars (% of participants indicating at least one)	42.00	35.70	55.20	57.10	50.00	50.00	58.20	50.00
Individual Education Plan (IEP; % of participants with one)	12.00	7.10	14.40	21.50	14.70	7.70	21.50	11.50
Number of Adults in the Home	1.78 (.71)	1.73 (0.60)	2.00 (.73)	1.98 (0.91)	2.08 (0.50)	2.11 (0.46)	2.20 (1.22)	2.21 (1.32)
Number of Children in the Home	1.44 (1.34)	1.19 (0.94)	1.65 (1.28)	1.76 (1.26)	1.21 (0.88)	1.11 (0.81)	1.73 (1.43)	1.72 (1.51)
Spanish Fluency task (*SD*)	7.88 (5.35)	7.78 (5.46)	2.54 (4.89)	2.11 (4.56)	12.86 (8.25)	14.48 (8.55)	3.98 (7.27)	1.95 (6.03)

Caregiver education level was coded as completed elementary school =1; completed middle school = 2; completed high school =3; some school beyond high school =4; completed an associate degree or other training program =5; completed a bachelor's degree =6; beyond college =7.

late elementary). Reflecting the diversity of the community, the sample was comprised of an approximately equal number of Black (n = 91), non-Hispanic White (n = 78), and Hispanic White (n = 92) participants, with an additional 27 participants identifying as multiracial. We recruited participants with letters distributed by their teachers and only those children whose parents/guardians provided written consent participated (52% of population; total providing consent n = 454). The reported sample (n= 288) reflects all students whose family consented and completed the parent/guardian questionnaire.

As described, the school system offers two educational tracks: mainstream English (monolingual) or two-way dual-language (TWDL; Spanish/English; 50% split instructional time). All children whose data were included in this study enrolled in their education model in kindergarten and maintained a stable placement. The TWDL program is housed within the school and TWDL and Mainstream English classrooms are alongside each other with the same resources. Entrance into the TWDL education model is by lottery at kindergarten registration. Not all families enter the lottery and the percentage of parents who do is not available. All children placed in the TWDL program are placed by lottery, but not all children in Mainstream English entered the lottery for TWDL placement. In light of the lack of complete randomization, we implemented Coarsened Exact Matching (CEM; Iacus, King, & Porro, 2012) to create a matched sample and examine differences in performance due to educational program assignment. This results in a quasi-experimental design (Shadish, Cook, & Campbell, 2002).

Measures

The task battery consisted of measures included for the purpose of creating matched-pairs, measures to examine group differences, and those that were the target of the investigation. For matching purposes, we included verbal and non-verbal intelligence as well as parent/guardian education level. Group difference measures included parent/guardian report of academic involvement and household density as well as a measure of Spanish fluency. Target variables included academic measures obtained from the participating school system and measures of executive functions, both computerized and teacher report.

Wechsler abbreviated scale of intelligence (WASI)

The WASI was designed to be a quick and reliable measure of general intelligence appropriate for ages 6–90, has reliabilities ranging from .92-.98, test–retest stability of .88 for children aged 6–11, and has been validated with other tests in the Weschler library (Weschler, 1999). We administered the Vocabulary subtest to measure word knowledge and verbal concept formation. We also administered the Block Design to measure nonverbal concept formation. Importantly, the WASI was not administered as a diagnostic tool and should not be interpreted as a clinical intelligence measure. This measure was administered for comparative purposes and was used as a matching variable. Both tests were conducted entirely in English. The total score was recorded for each subtest, resulting in two independent variables, which were used to create the matched sample.

Parent and teacher data

Parents completed a written questionnaire including family demographics and students' activities (See Table 2). With parental authorization, the school guidance counselors provided academic data for participating children (see Table 2). The measures collected differed by grade. In primary (K and 1st grade), language arts (reading and writing) performance was measured with the mCLASS: Text Reading Comprehension (TRC). The TRC tracks reading development across the elementary years. Children read leveled text passages and complete comprehension questions. Expected progress is a level D, or level 6, by the end of kindergarten and a level J, or 18, by the end of first grade. The math measure was year-end-grade achieved by each student as an accumulation of classroom assessments across the academic year. For intermediate (4th and 5th grades), we collected year-end grades for language arts, math, science, and social studies for a total of four means. Grades are on a 100 point

Table 2. Mean scores (and *SD*s) on tests of academic achievement by grade level and educational model.

	Primary				Intermediate			
	Dual-Language		Mainstream		Dual-Language		Mainstream	
Academic Means	Full	CEM	Full	CEM	Full	CEM	Full	CEM
Language Arts (*SD*)								
Kindergarten TRC	6.89 (3.43)	7.35 (3.25)	5.88 (3.43)	5.80 (3.47)				
First Grade TRC	16.46 (5.64)	16.33 (4.56)	19.02 (6.37)	20.22 (5.82)				
Year-average					93.24 (3.59)	93.72 (2.87)	88.05 (6.49)	90.16 (5.28)
End-of-grade					2.15 (0.91)	2.08 (0.81)	1.81 (0.83)	2.08 (0.86)
Math (*SD*)								
Kindergarten Year-average	84.64 (9.35)	84.18 (9.22)	82.71 (13.99)	82.13 (14.10)				
First Grade Year = average	80.08 (11.39)	81.39 (11.67)	83.77 (12.83)	82.17 (15.70				
Year-average					94.15 (3.29)	94.36 (2.72)	85.91 (7.50)	87.88 (6.90)
End-of-grade					2.41 (0.89)	2.42 (0.90)	1.80 (0.96)	1.96 (0.98)
Science (*SD*)								
Year-average					94.73 (1.86)	94.72 (2.01)	90.74 (5.05)	92.40 (3.42)
Social Studies (*SD*)								
Year-average					96.15 (2.99)	96.56 (2.65)	90.65 (6.32)	91.32 (5.13)

TRC is Text Reading and Comprehension. Year-average grades refer to those determined by classroom teachers based on school-wide assessments. End-of-grade refers to standardized state tests.

scale. We also collected scores on the state-standardized end-of-grade tests (EOG). The tests are administered beginning in the third grade in all public schools in North Carolina. Scores range from 1 to 4, with a three the minimum to advance to the next grade without remediation. Both 4th and 5th grade students take the language arts and math tests, for a total of 2 means.

Spanish category fluency

Although time restrictions from the collaborating school system prevented a full language test battery, the fluency task allowed us to get an indication of whether children in the TWDL model were developing Spanish language skills at a higher rate than their peers in the mainstream model, an important premise of bilingual education and the bilingual advantage. Tasks such as these, although administered in English, are often part of intelligence scales (such as the *Weschler Intelligence Scale for Children, Revised*; Weschler, 1974). In this task children were first asked to name as many animals within a minute as they could, using Spanish, "such as perro [dog] or gato [cat]" and then to name as many things to eat or drink as they could within one minute, using Spanish, "such as leche [milk] or pan [bread]." Experimenters recorded correct, unique, responses and the total number was summed for one measure of Spanish Category Fluency.

Computerized executive functions measures

Computerized tasks used the Psychology Experiment Building Language (PEBL: Mueller, 2011) delivered on Compaq Presario CQ60 laptop computers attached to HP Compaq L2105 21.5 inch color touch screen monitors.

Switching. We administered a version of the Trail Making Task developed specifically for children (Delis, Kaplan, & Krames, 2001). Children completed three sequencing trails. The first trail required numeration of numbers from 1 to 16 (1-2-3). In the second, the sequence was alphabetical from A-K (A-B-C). In the last, the sequence alternated between a number and a letter (1-A-2-B-3-C).

Response time in seconds is calculated for each trail, with the last trail being a measure of switching performance. The PEBL version of this task has been tested for validity (Piper et al., 2012) and higher test–retest reliability than paper versions (*r*= .61-.74 vs. *r*= .45, respectively; Piper et al., 2015). Bilingual children have previously shown an advantage over monolingual children on this task (Bialystok, 2010; Esposito & Baker-Ward, 2013).

Inhibitory control. We included both a stimulus-stimulus conflict measure as well as a stimulus-response conflict measure of inhibitory control. The Bivalent Shape task (stimulus-stimulus conflict: Mueller & Esposito, 2014) was developed to measure the ability to ignore salient features of a stimulus and act on the relevant information. The task has been used to measure executive functions in this age range previously (e.g., Esposito & Bauer, 2018). Two active buttons are at the bottom of the screen, a red circle, and a blue square. Stimuli (circles and squares in red, blue, or a black outline for six possible test items) appeared in the center of the screen and participants are directed to match the shape. Congruent stimuli matched in both color and shape, incongruent stimuli matched in shape but not in color. Neutral items did not have color, and thus no facilitating or distracting element. The task consisted of practice blocks for each type of stimuli followed by a mixed block in which the three types of stimuli were presented in a set randomized order with 10 of each type for a total of 30 trials. Children had up to 3 seconds to respond. Mean reaction times on correct trials for each trial type produce three dependent variables (congruent, incongruent, and neutral). Bilingual children have shown an advantage over their monolingual peers on this task (Esposito, Baker-Ward, & Mueller, 2013).

The Simon task represents a stimulus-response conflict. Participants responded to the appearance of a green or orange rectangle on the left or right side of the monitor by pressing the right or left shift buttons in response to the color, ignoring the location of the object on the screen (Lu & Proctor, 1995; Simon & Wolf, 1963). The task has good reliability (Cronbach's alpha = .88, with adults in short form; Cevada, Conde, Marques, & Deslandes, 2019) and has been validated as a marker of attention deficit along with other executive function tasks in children (e.g., Mullane, Corkum, Klein, & McLaughlin, 2009). Individuals respond more quickly if the object is on the same side of the screen as the shift button that correlates to the color and they respond more slowly when the position of the object is on the opposite side of the appropriate shift button. The Simon task has a history of elucidating a bilingual advantage, especially in children (e.g., Martin-Rhee & Bialystok, 2008). Mean reaction times for correct trials were calculated for congruent and incongruent trial types for a total of two dependent variables.

Behavioral executive functions measure

Teachers completed the Behavioral Rating Inventory of Executive Functions (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2015). The BRIEF is an 86 item behavioral checklist normed for children aged 5–18. Respondents mark each behavior as "never," "sometimes," or "always" and the responses are summed into a Global Executive Composite that documents executive dysfunction. Reliability as measured by Cronbach's alpha ranges from .8-.98. The test is validated across age, socio-economic, race, and ethnic groups within the United States. A higher score indicates greater dysfunction.

Procedures

Children were tested individually in a quiet classroom within their school during a single session lasting approximately 30 minutes (the time allotted by the participating school system). Experimenters were six female psychology students with intermediate to advanced Spanish ability. All participating children provided verbal assent. The university institutional review board and participating school system school board reviewed and approved all procedures. The computer tasks were administered in random order. All computer tasks are designed to be non-verbal and instructions were administered in the child's preferred language (children were asked their preference; English or Spanish). Computer tasks were followed by the Spanish fluency task. We concluded with the WASI, which was administered in English as per the manual. All assessments, reports, and teacher ratings were completed in the last month of the school year.

Matching

In order to control for individual and group differences that could influence the outcome variables of interest as well as control for differences in sample size between groups, we created a matched-sample using Coarsened Exact Matching. CEM analysis utilized an R plugin (R version 3.3.0; CEM Extension Bundle). Each TWDL participant was paired to a child enrolled in mainstream English education. Grade was exactly matched while verbal intelligence, non-verbal intelligence, and parent/guardian education level were coarsened. Of the 288 participants, 68 pairs were created for a total of 136 participants included in the CEM subset. Means are reported in Table 1, with matching variables outlined for easy identification and separated by Education Model (TWDL vs. Mainstream English) and School (primary vs. intermediate).

Results

The results are reported in three parts; examination of group differences, academic performance, and executive functions. Evidence for executive functions as a mediator for the relation between education model and academic performance in then examined. All reported analyses are two-tailed and were conducted using SPSS 24 software.

Group differences

We first examined whether the educational groups differed in observable ways that could impact academic performance (means reported by group in Table 1). Group differences in verbal and non-verbal intelligence, parent/guardian level of education, parent/guardian academic involvement (homework help and reading), participation in extra-curricular activities, child has an individual educational plan (IEP; implemented for special needs), and the density of the home environment were tested with a 2 (Educational Model) x 2 (School) multivariate analyses of variance (MANOVA). There was a main effect of School, $F(9, 185) = 24.42$, $p < .001$, $\eta^2 = .54$, such that the intermediate students, compared to the primary students, scored higher in verbal ($M = 19.01$ vs. 36.63; $F(1, 197) = 160.44$, $p < .001$, $\eta^2 = .45$) and non-verbal intelligence ($M = 7.75$ vs. 19.59; $F(1, 197) = 85.24$, $p < .001$, $\eta^2 = .31$). Families with children in the intermediate school, compared to the primary school, reported spending fewer hours supervising homework ($M = 2.44$ vs. 3.36; $F(1, 197) = 5.56$, $p = .02$, $\eta^2 = .03$) and reading with their child ($M = 1.98$ vs. 3.12; $F(1, 197) = 8.81$, $p = .003$, $\eta^2 = .04$). There were no other main effects or interactions.

The CEM procedure was intended to minimize differences between education groups on socio-economic and general aptitude. However, we recognize that the measures included are a proxy and incomplete. We, thus, examined whether the educational groups differed in observable ways that could impact academic performance but were not included in the matching variables in a 2 (Education Model) x 2 (School) MANOVA with parent/guardian academic involvement (homework help and reading), participation in extra-curricular activities, child has an IEP, and the density of the home environment as dependent variables. The results were similar to those of the full-sample analyses. There was a main effect of School, $F(6, 96) = 2.72$, $p = .02$, $\eta^2 = .15$, such that families in the intermediate school reported spending less time supervising homework ($M = 2.51$ vs. 3.36; $F(1, 105) = 5.71$, $p = .02$, $\eta^2 = .05$) and reading with their children ($M = 2.01$ vs. 3.31; $F(1, 105) = 12.44$, $p = .001$, $\eta^2 = .11$). There was also a main effect of Education Model, $F(6, 96) = 2.26$, $p = .04$, $\eta^2 = .12$, such that families in the dual-language program reported fewer children in the home compared to those in mainstream education (M = 1.16 vs. 1.74; $F(1, 105) = 4.77$, $p = .03$, $\eta^2 = .05$). There were no other main effects and no interactions.

We also examined the Spanish Fluency of participants. We examined this in a 2 (Education Model) x 2 (School) Analysis of Variance (ANOVA).[1] In the full sample, there was a main effect of Education Model such that students in the TWDL program had higher scores than their peers in the mainstream model, $F(1, 238) = 65.74$, $p < .001$, $\eta^2 = .22$. There was a main effect of School such that the intermediate students had higher performance compared to the primary students, $F(1, 238) = 13.44$, $p < .001$, $\eta^2 = .05$. There was also a significant interaction, $F(1, 238) = 4.08$, $p = .04$, $\eta^2 = .02$. Follow-up univariate tests revealed that the dual-language intermediate students had significantly higher performance than the primary students, $F(1, 77) = 10.49$, $p = .002$, whereas there was no difference between educational programs for those in the mainstream education program, $F(1, 159) = 2.25$, $p = .14$.

The same analyses with the CEM sample replicated these results. There were significant main effects of Education Model, $F(1, 119) = 61.86$, $p < .001$, $\eta^2 = .35$, and School, $F(1, 119) = 7.98$, $p = .006$, $\eta^2 = .07$, and a significant interaction, $F(1, 119) = 8.77$, $p = .004$, $\eta^2 = .07$. Follow-up univariate tests also replicated the previous analyses with dual-language intermediate students producing significantly more words than primary students, $F(1, 62) = 14.06$, $p < .001$, $\eta^2 = .19$, while there was no difference for those in mainstream education, $F(1, 57) = 0.01$, $p = .91$.

In summary, there were few differences found between the Education models with the exception of Spanish fluency performance. There were also few differences between the full sample and the CEM sample.

Academic performance

Primary school

Academic performance is reported by grade and education model in Table 2.[2] Primary school academic achievement was represented with a reading measure (TRC) and math measure (year-end classroom grade). For the primary school, we had access to kindergarten data for all participants (including current first-grade students). Thus, we were able to analyze kindergarten data for all primary students (n = 156) as a MANOVA with Education Model as a predictor. The MANOVA revealed no significant differences in reading or mathematical performance between Education Models, $F(2, 153) = 1.43$, $p = .24$. In addition to kindergarten data, first-grade students also had data from their first-grade year. This enabled a short-term longitudinal examination of the growth between kindergarten and first grade for children in the TWDL program compared to those in the mainstream education model. We did a repeated measures MANOVA with Education Model as a predictor. There was a main effect such that math scores decreased between kindergarten and first grade (M = 1.98 vs. 3.12; $F(1, 72) = 55942.04$, $p < .001$, $\eta^2 = .99$). It is important to note that the math measure is based on the content from that year, so a lower grade does not indicate *less knowledge* overall, but poorer performance on the more advanced content. There were no other significant main effects or interactions. The interaction between reading and Education Model approached significance ($p = .08$) such that children in the TWDL program showed less growth in reading compared to their peers in mainstream education.

Intermediate school

Intermediate school academic achievement was represented with both standardized test scores (EOG) in mathematics and language arts and classroom grades in mathematics, language arts, science, and social studies. Both 4th and 5th grade students completed these measures and the measures are normed within grade, meaning we would not expect to see growth across grade levels. We analyzed the standardized test scores in a MANOVA with Education Model as a predictor. Education Model was significant, $F(2, 105) = 4.53$, $p = .01$, $\eta^2 = .08$. Univariate analyses revealed that students in the TWDL program had significantly higher performance on the standardized math score, $F(2, 108) = 9.13$, $p = .003$, $\eta^2 = .08$, and the standardized language arts score neared significance in the same direction, $F(2,$

108) = 3.58, p= .06, η^2 = .03. A MANOVA predicting classroom grades in math, language arts, science, and social studies with educational program as the predictor revealed similar results. Educational program was significant, $F(4, 102) = 9.26$, $p < .001$, $\eta^2 = .27$, and univariate analyses revealed higher performance for TWDL students in all subjects, $Fs(1, 107) \geq 362.25$, $ps < .001$, $\eta^2s \geq .15$.

Primary school CEM

We used the same plan of analyses to examine academic performance within the CEM sample. A MANOVA examining kindergarten reading and math performance replicated the full sample in that there was not a significant difference by Education Model, $F(2, 75) = 1.70$, $p = .19$. The repeated measures MANOVA examining kindergarten and first-grade reading and math performance predicted by Education Model also replicated the findings in the full sample. There was a main effect of math such that math scores decreased between kindergarten and first grade, $F(1, 28) = 2891.51$, $p < .001$, $\eta^2 = .99$, but no other significant effects.

Intermediate school CEM

In the MANOVA examining intermediate student standardized math and language arts performance, Education Model was not a significant predictor of performance, $F(2, 47) = 1.93$, p= .12. The MANOVA predicting classroom grades in math, language arts, science, and social studies with Educational Program as the predictor replicated results with the full sample. Educational program was significant, $F(4, 45) = 6.11$, $p = .001$, $\eta^2 = .35$, and univariate analyses revealed higher performance for TWDL students in all subject areas, $Fs(1, 50) \geq 8.57$, $ps \leq .005$, $\eta^2s \geq .15$.

In summation, there were no significant differences between educational programs in performance for primary students in either the full or CEM samples. Intermediate students in the TWDL program significantly outperformed their mainstream-educated peers on the standardized math exam and on classroom grades in all subjects in the full sample. In the CEM sample, students in TWDL education had significantly higher scores in classroom grades.

Executive functions

Executive functions were measures both with computerized tasks as well as a teacher reported inventory. Scores on executive functions tasks are shown in Table 3. Preliminary analyses indicated that the executive functions tasks (TMT, BST, Simon, BRIEF) were measuring unique attributes of executive functions as intended, $rs < .46$.[3] Therefore, we ran separate analyses for the outcome variables of each measure: TMT (Trail A, Trail B, Trail C), BST (neutral, congruent, incongruent), and Simon (congruent and incongruent), and BRIEF. In all four,[4] we utilized 2 (Education Model) x 2 (School) MANOVAs or an ANOVA (for BRIEF) to test for main effects of Education Model and School as well as an interaction. To ensure that age did not have an undue influence on the models, we calculated Z-scores ($M = 0$, $SD = 1$) for each of the computerized executive function measures within each grade. The conversion to Z-scores allowed us to examine relative ranking within each grade. The BRIEF is measures based on grade-level expectations and did not require Z-score conversion.

Full sample

We first analyzed the data from the full sample. There were no significant main effects or interactions for the TMT, $Fs(3, 237) \leq 2.45$, $ps \geq .06$, or the BST, $Fs(3, 233) \leq 2.36$, $ps \geq .07$. There were no significant main effects on the Simon task, $Fs(2, 236) \leq 1.40$, $ps \geq .25$, but there was a significant interaction, $F(2, 236) = 3.48$, $p = .03$, $\eta^2 = .03$. Univariate analyses revealed that intermediate TWDL students were significantly faster in the congruent trials compared to their mainstream educated peers, $F(1, 241) = 6.90$, $p = .009$, $\eta^2 = .03$. Thus, though the means indicate the intermediate students were faster in the computerized executive functions tasks, this difference only reached significance on one part of one task.

Table 3. Mean reaction times in milliseconds for executive functions tasks by school, trial type, and education model.

		Primary				Intermediate			
		Dual-Language		Mainstream		Dual-Language		Mainstream	
Task	Trial Type	Full	CEM	Full	CEM	Full	CEM	Full	CEM
BST (*SD*)									
	Neutral	1.37 (0.22)	1.36 (0.23)	1.37 (0.21)	1.38 (0.24)	0.98 (0.12)	0.96 (0.12)	1.08 (0.14)	1.08 (0.14)
	Congruent	1.40 (0.23)	1.41 (0.23)	1.35 (0.23)	1.36 (0.24)	0.98 (0.14)	0.97 (0.15)	1.03 (0.12)	1.03 (0.11)
	Incongruent	1.43 (0.23)	1.42 (0.23)	1.43 (0.26)	1.48 (0.27)	1.04 (0.17)	1.03 (0.17)	1.10 (0.16)	1.11 (0.16)
TMT (*SD*)									
	Trail A	27.17 (11.23)	27.96 (11.79)	23.75 (12.00)	26.93 (15.19)	11.41 (4.72)	10.80 (3.50)	13.58 (7.26)	12.46 (4.20)
	Trail B	27.81 (16.04)	28.79 (16.98)	27.14 (15.23)	29.70 (15.01)	13.94 (6.51)	12.68 (3.66)	15.71 (11.66)	13.30 (3.78)
	Trail C	52.08 (24.65)	53.35 (25.64)	53.76 (31.85)	58.63 (42.70)	24.98 (11.01)	25.27 (9.90)	31.42 (15.45)	27.93 (16.74)
Simon (*SD*)									
	Congruent	1.52 (0.44)	1.57 (0.45)	1.41 (0.30)	1.44 (0.34)	0.98 (0.11)	0.95 (0.11)	1.05 (0.14)	1.06 (0.13)
	Incongruent	1.60 (0.42)	1.62 (0.44)	1.56 (0.34)	1.63 (0.40)	1.05 (0.13)	1.04 (0.12)	1.13 (.15)	1.11 (0.17)
BRIEF GEC		89.72 (21.75)	88.85 (19.05)	102.78 (31.12)	104.50 (31.63)	82.67 (17.87)	81.64 (18.92)	101.35 (31.87)	101.00 (27.09)

BST refers to Bivalent Shape Task. TMT refers to Trail Making Task. GEC refers to Global Executive Composite.

In a 2 (Education Model) x 2 (School) ANOVA predicting the Global Executive Composite score, there was a main effect of Education Model such that children in the TWDL program exhibited fewer disordered behaviors related to executive functions, $F(1, 259) = 15.44$, $p < .001$, $\eta^2 = .06$. There was not a main effect of School, $F(1, 259) = 1.10$, $p = .30$, and no interaction, $F(1, 259) = 0.48$, $p = .49$.

CEM sample

We next analyzed the data from the CEM sample. Replicating the full sample, there were no significant main effects or interactions for the TMT, $Fs(3, 116) \leq 0.96$, $ps \geq .04$. In contrast, there was a significant main effect of Education Model in the BST, $F(3, 113) = 2.79$, $p = .04$, $\eta^2 = .07$, such that children in the TWDL program were faster at responding to neutral trials, $F(1, 115) = 6.37$, $p = .01$, $\eta^2 = .05$, and trending in that direction for incongruent trials, $F(1, 115) = 3.75$, $p = .05$, $\eta^2 = .03$, compared to their mainstream educated peers. There were no other main effects or interactions for the BST, $Fs(3, 113) \leq 1.55$, $ps \geq .21$. The Simon task analyses again replicated the results of the full sample in that there were no significant main effects, $Fs(2, 117) \leq 0.86$, $ps \geq .43$, but there was a significant interaction, $F(2, 117) = 4.60$, $p = .01$, $\eta^2 = .07$. Univariate analyses revealed that intermediate TWDL students were significantly faster in the congruent trials compared to their mainstream peers, $F(1, 122) = 7.57$, $p = .007$, $\eta^2 = .06$.

In a 2 (Education Model) x 2 (School) ANOVA predicting the Global Executive Composite score on the BRIEF, the results replicated those of the full sample. There was a main effect of Education Model such that children in the TWDL program exhibited fewer disordered behaviors, $F(1, 120) = 13.88$, $p < .001$, $\eta^2 = .11$. There was not a main effect of School, $F(1, 120) = 1.30$, $p = .26$, and no interaction, $F(1, 120) = 0.16$, $p = .69$.

In summary, the computerized executive function tasks did not reveal a robust pattern of results. The BRIEF, in contrast, indicated a main effect such that the students in the TWDL program were exhibiting behaviors consistent with more developed executive functions compared to their peers in the mainstream education model. This pattern of results was found in both the full sample and the CEM sample.

Table 4. Regression models analyzing the potential of executive functions as a mediator for education model and math achievement.

Model	*n*	*t*	*p*	β	*F*	*df*	*p*	*adj.* R^2
Step 1; DV = Executive Functions	101				9.81	1, 99	.002	.08
Educational Program		−3.13	.002	−.30				
Step 2; DV = Math performance	101				8.33	1, 99	.005	.08
Educational Program		2.89	.005	.28				
Step 3; DV = Math performance	101				14.83	1, 99	< .001	.12
Executive Functions		−3.85	< .001	−.36				
Step 4; DV = Math performance	101				9.48	2, 98	< .001	.15
Educational Program		1.93	.06	.19				
Executive Functions		−3.14	.002	−.31				

Mediation analyses

We next explored whether executive functions is a potential mechanism through which intermediate TWDL students are showing an academic advantage over their mainstream educated peers. Specifically, we examined whether the teacher reported BRIEF mediated the relation between being in the TWDL program and academic performance on the standardized measure of math achievement. This analysis was only conducted with the standardized math measure in the full sample for three reasons. First, classroom teachers issue both classroom grades and the behavioral rating. Any relation between the variables could be due to coming from the same source. Second, only math was assessed because the standardized measure of language arts did not significantly differ between Education Models. Third, the significant difference in standardized math performance was only found in the full sample.

In this analysis, we followed the four steps outlined by Baron and Kenny (1986) to assess mediation (see Table 4 for regression results). First, we examined whether Education Model significantly predicted executive functions. Second, we conducted a regression with Education Model predicting standardized math performance. Third, we conducted a regression with executive functions predicting math performance. Fourth, we examined whether the effect of Education Model on math performance was reduced when executive functions was included in the regression. Having determined mediation with these steps, we next examined whether the mediation was statistically significant with the Sobel test (1982). The Sobel test indicated statistically significant mediation, $z = 2.47$, $p = .01$.

Discussion

The present study investigated the academic achievement, executive functions, and the relation between these variables in a cross-sectional design comparing students in primary or intermediate elementary education in either a two-way dual-language program or a mainstream English education program. The pattern of results supported an academic advantage for intermediate TWDL students. The advantage in executive functions was less robust, emerging for TWDL students in behavioral ratings but not in computerized measures. Using the behavioral rating measure of executive functions and a standardized measure of math performance, we did find evidence for executive functions as a mechanism supporting the academic advantage.

We predicted an academic advantage for intermediate students in the TWDL program, who had experienced the program for 5–6 years. As predicted, no differences were found in academic performance for primary students. Intermediate TWDL students showed an advantage in both standardized measures of achievement and classroom grades in the full-sample. The match-sample showed a similar pattern, but the standardized measures failed to reach significance. Classroom grades had a greater range of performance compared to the narrow range of the standardized measures. This limited variability could contribute to the differences in the results between the full sample that included a larger sample and did not reduce participants on the extreme ends of the distribution (in comparison to the matched sample).

While there were few differences between those in TWDL and mainstream education in the computerized measures of executive functions, the behavioral rating measure revealed a significant difference between education models such that children in the TWDL program exhibited fewer indicators of executive dysfunction in the classroom. The difference was present at both the primary and the intermediate level. This could indicate a preexisting group difference such that children in the TWDL program begin school with greater executive functions skills. The results could also indicate that teachers in the TWDL program have different expectations of students in regards to classroom behavior and, thus, rate them more favorably. There could also be other group differences aside from education program that our variables failed to capture. Alternatively, it could indicate that an advantage in executive functions related to classroom behavior emerges early after only one or two years of participation in a TWDL program. There are at least two indicators of the latter interpretation. First, teachers in the TWDL program and the mainstream program are within the same community with the same training opportunities and regularly rotate through teaching in the TWDL program. Thus, differences in expected behaviors to the degree found in this study are unlikely. Second, the differences in the scores predicted differences in academic achievement on a state standardized test. This last point indicates that the differences in performance were not a systematic bias held by teachers and were instead meaningful reflections of behavior predicting academic outcomes.

We analyzed whether executive functions is a potential mechanism for the academic advantage often found for children in dual-language education models. The analysis was limited to intermediate students who had a standardized measure of academic performance not issued by a teacher. This eliminated the concern that both measures of interest were coming from the same source or any concern regarding different expectations across programs. The results did indicate that the academic advantage found on the standardized math assessment for children at the intermediate level of the TWDL program was mediated by executive functions behaviors exhibited in the classroom. Although the analyses were limited to a subset of the participants and variables, the results support further investigation of cognitive advantages emerging through TWDL education that could positively influence academic achievement. Bilingual education is generally focused on bilingualism, biliteracy, and cultural competence, but the present results are the first evidence we are aware of that children in a bilingual education program have advanced executive skills that positively impact their academic achievement. Further research is needed to determine whether executive functions are developing differently over time based on educational program placement, but these results are encouraging.

The executive functions advantage for TWDL students was present across all grade levels, yet the academic advantage was only present in the intermediate students. If we assume that the executive functions advantage is not the results of preexisting group differences, the results indicate that, in comparison, it takes longer for the academic advantage to emerge. This is possibly because executive functions behaviors supporting the academic advantage do not have an immediate effect. It could also indicate that a language threshold needs to be met in both languages before the academic advantage emerges. Future research that includes full language assessments could help to understand the relation between language development, academic achievement, and executive functions in dual-language education models.

An interesting finding was the difference between the results of the computerized executive functions measures and the behavioral ratings. Studies of executive functions, including those used to identify the bilingual advantage and the contexts in which it emerges, rely heavily on computerized measures (for a review, see Bialystok et al., 2012). The computerized measures, in theory, are measuring abilities utilized in the classroom. However, performance on computerized tasks has not shown good transfer to how children are actually behaving in classrooms. Children who are trained on a computerized tasks in laboratories show improvements on that task and sometimes similar computerized tasks (near transfer), but rarely show benefits in areas such as academic performance (considered far transfer; Kassai, Futo, Demetrovics, & Takacs, 2019; Serpell & Esposito, 2016). The results of the present study indicate that advantages in executive functions gained through bilingual education may be more readily detected with an applied measure relevant to classroom behavior.

A recurring criticism of work identifying the bilingual advantage in executive functions is that it may be based on differences in SES (see Valian, 2015, for review). A similar concern has arisen regarding the academic advantage for children in dual-language education. The explanation is that lower SES students, who are more likely to be transient, most commonly attrit out of the program, leaving higher SES students in the later elementary years where the academic advantage is most often found (e.g., Hill, 2018). As a way of addressing these concerns, we analyzed the data twice. First, we included all consenting participants who completed the tasks of interests. In this full sample, we examined group differences in participants' intelligence, parent education level, and other indicators, but we did not control for these measures. We then analyzed a subset of the data that was matched on grade level (exact) as well as verbal and non-verbal intelligence and parent education level (coarsened exact matching). The intention was to identify whether group differences found in the full sample would still be found when controlling for individual and family-level variables known to affect academic outcomes and executive functions. Although there were some small changes in the results regarding which analyses reached significance, the overall pattern between the two samples was similar. The results do not support a stance that advantages in academic achievement or executive functions are the result of group differences in intelligence or SES, nor do they support attrition as an explanation for later emerging academic advantages.

The present study is not without limitations. The study is a cross-sectional and quasi-experimental investigation that captures a snapshot of the student performance. A longitudinal investigation that included random assignment to educational program and measurement prior to school start would be better able to capture the developmental changes attributed to the educational model. However, the school system allows parents to opt-in, eliminating random assignment as a possibility. We were also limited in time with each child by the participating school system and, thus, were unable to assess language proficiency. As future directions, we recommend longitudinal work that will be able to assess not only a snapshot of current performance, but the change over time as cross-sectional research cannot identify preexisting group differences. The current work is also focused specifically on an English-only model (teachers do not have proficiency in other languages) compared to a dual-language model. Additionally, we recommend assessments of language proficiency in both the partner languages as well as translanguaging practices (García & Lin, 2016) to identify the role of language development in both academic performance and executive functions for both education models.

In conclusion, the present study found evidence for executive functions as a mechanism contributing to the academic advantage shown by TWDL intermediate students compared to their mainstream English education peers. The present study indicates that in addition to gaining second-language fluency, literacy, and greater cultural competence, two-way dual-language education models contribute to cognitive development, specifically executive functions, in ways that support academic achievement.

Notes

1. Analyses, including in CEM models, were run with and without a variable reflecting home language as reported by parents. The variable did not change pattern of results. The more parsimonious model is, therefore, reported.
2. Analyses were run with and without a variable reflecting English proficiency at kindergarten entry. Results did not differ between models. Therefore, the more parsimonious model is reported.
3. Latent variable analyses was attempted, but the model did not converge.
4. Analyses were run with and without a variable reflecting English proficiency at kindergarten entry. Results did not differ between models. Therefore, the more parsimonious model is reported.

Funding

This work was supported by a postdoctoral/predoctoral fellowship provided by the National Institute of Child Health and Human Development (T32-HD07376) through the Center for Developmental Science, University of North Carolina at Chapel Hill, and by a subaward of IES R305A150492 to Alena G. Esposito.

References

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. doi:10.1037/0022-3514.51.6.1173

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences*, *21*(4), 327–336. doi:10.1016/j.lindif.2011.01.007

Bialystok, E. (2010). Global–local and trail-making tasks by monolingual and bilingual children: Beyond inhibition. *Developmental Psychology*, *46*(1), 93. doi:10.1037/a0015466

Bialystok, E., & Barac, R. (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition*, *122*(1), 67–73. doi:10.1016/j.cognition.2011.08.003

Bialystok, E., Craik, F. I., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240–250. doi:10.1016/j.tics.2012.03.001

Cevada, T., Conde, E., Marques, D., & Deslandes, A. C. (2019). Test-retest reliability of the simon task: A short version proposal. *Somatosensory & Motor Research*, *36*(4), 275–282. doi:10.1080/08990220.2019.1689114

Cobb, B., Vega, D., & Kronauge, C. (2006). Effects of an elementary dual language immersion school program on junior high school achievement. *Middle Grades Research Journal*, *1*(1).

De Cat, C., Gusnanto, A., & Serratrice, L. (2018). Identifying a threshold for the executive function advantage in bilingual children. *Studies in Second Language Acquisition*, *40*(1), 119–151. doi:10.1017/S0272263116000486

Delis, D. C., Kaplan, E., & Krames, J. H. (2001). *Trail making task. Delis Kaplan executive function system examiner manual* (pp. 39–54). San Antonio, Texas: The Psychological Corporation Harcourt Assessment Company.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135–168. doi:10.1146/annurev-psych-113011-143750

Esposito, A. G., & Baker-Ward, L. (2013). Dual-language education for low-income children: Preliminary evidence of benefits for executive function. *Bilingual Research Journal*, *36*(3), 295–310. doi:10.1080/15235882.2013.837848

Esposito, A. G., Baker-Ward, L., & Mueller, S. T. (2013). Interference suppression vs. response inhibition: An explanation for the absence of a bilingual advantage in preschoolers' Stroop task performance. *Cognitive Development*, *28*(4), 354–363. doi:10.1016/j.cogdev.2013.09.002

Esposito, A. G., & Bauer, P. J. (2018). Building a knowledge base: Predicting self-derivation through integration in 6-to10-year-olds. *Journal of Experimental Child Psychology*, *176*, 55–72. doi:10.1016/j.jecp.2018.07.011

García, E. E., & Náñez Sr, J. E. (2011). *Bilingualism and cognition*. American Psychological Association.

García, O., & Lin, A. M. (2016). Translanguaging in bilingual education. In O. García, A. Lin, & S. May (Eds.), *Bilingual and multilingual education* (pp. 117–130), Encyclopedia of Language and Education 5. New York: Springer.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2015). *BRIEF: Behavior rating inventory of executive function*. Lutz, FL: Psychological Assessment Resources.

Hartanto, A., Toh, W. X., & Yang, H. (2018). Bilingualism narrows socioeconomic disparities in executive functions and self-regulatory behaviors during early childhood: Evidence from the early childhood longitudinal study. *Child Development*, *90*(4), 1215–1235. doi:10.1111/cdev.13032

Hill, S. R. (2018). A meta-analysis of the effect of ten language immersion programs on academic achievement. *Spanish and Portuguese Review*, *4*, 23–50.

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology*, *49*(1), 4. doi:10.1037/a0027238

Howard, E. R., Lindholm-Leary, K. J., Rogers, D., Olague, N., Medina, J., Kennedy, B., … Christian, D. (2018). *Guiding principles for dual language education* (3rd ed.). Washington, DC: Center for Applied Linguistics.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, *20*(1), 1–24. doi:10.1093/pan/mpr013

Janus, M., Lee, Y., Moreno, S., & Bialystok, E. (2016). Effects of short-term music and second-language training on executive control. *Journal of Experimental Child Psychology*, *144*, 84–97.

Kassai, R., Futo, J., Demetrovics, Z., & Takacs, Z. K. (2019). A meta-analysis of the experimental evidence on the near- and far-transfer effects among children's executive function skills. *Psychological Bulletin*, *145*(2), 165. doi:10.1037/bul0000180

Kaushanskaya, M., Gross, M., & Buac, M. (2014). Effects of classroom bilingualism on task-shifting, verbal memory, and word learning in children. *Developmental Science*, *17*(4), 564–583. doi:10.1111/desc.12142

Lindholm-Leary, K., & Genesee, F. (2014). Student outcomes in one-way, two-way, and indigenous language immersion education. *Journal of Immersion and Content-Based Language Education*, *2*(2), 165–180. doi:10.1075/jicb.2.2.01lin

Lu, C. H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review*, *2*(2), 174–207. doi:10.3758/BF03210959

Marian, V., Shook, A., & Schroeder, S. R. (2013). Bilingual two-way immersion programs benefit academic achievement. *Bilingual Research Journal*, *36*(2), 167–186. doi:10.1080/15235882.2013.818075

Martin-Rhee, M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *11*(1), 81–93. doi:10.1017/S1366728907003227
Mueller, S. T. (2011). *The psychology experiment building language (Version 0.11)* [Computer software]. Retrieved from http://pebl.sourceforge.net
Mueller, S. T., & Esposito, A. G. (2014). Computerized testing software for assessing interference suppression in children and adults: The Bivalent Shape Task (BST). *Journal of Open Research Software*, *2*(1), e3.
Mullane, J. C., Corkum, P. V., Klein, R. M., & McLaughlin, E. (2009). Interference control in children with and without ADHD: A systematic review of Flanker and Simon task performance. *Child Neuropsychology*, *15*(4), 321–342. doi:10.1080/09297040802348028
Nesbitt, K. T., Baker-Ward, L., & Willoughby, M. T. (2013). Executive function mediates socio-economic and racial differences in early academic achievement. *Early Childhood Research Quarterly*, *28*(4), 774–783. doi:10.1016/j.ecresq.2013.07.005
Nicolay, A. C., & Poncelet, M. (2013). Cognitive advantage in children enrolled in a second-language immersion elementary school program for three years. *Bilingualism: Language and Cognition*, *16*(3), 597–607. doi:10.1017/S1366728912000375
Nicolay, A. C., & Poncelet, M. (2015). Cognitive benefits in children enrolled in an early bilingual immersion school: A follow up study. *Bilingualism: Language and Cognition*, *18*(4), 789–795. doi:10.1017/S1366728914000868
Padilla, A. M., Fan, L., Xu, X., & Silva, D. (2013). A Mandarin/English two-way immersion program: Language proficiency and academic achievement. *Foreign Language Annals*, *46*(4), 661–679. doi:10.1111/flan.12060
Piper, B. J., Li, V., Eiwaz, M. A., Kobel, Y. V., Benice, T. S., Chu, A. M., . . . Mueller, S. T. (2012). Executive function on the psychology experiment building language tests. *Behavior Research Methods*, *44*(1), 110–123. doi:10.3758/s13428-011-0096-6
Piper, B. J., Mueller, S. T., Geerken, A. R., Dixon, K. L., Kroliczak, G., Olsen, R. H., & Miller, J. K. (2015). Reliability and validity of neurobehavioral function on the Psychology Experimental Building Language test battery in young adults. *PeerJ*, *3*, e1460. doi:10.7717/peerj.1460
Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, *113*(4), 535–551. doi:10.1016/j.jecp.2012.06.013
Purić, D., Vuksanović, J., & Chondrogianni, V. (2017). Cognitive advantages of immersion education after 1 year: Effects of amount of exposure. *Journal of Experimental Child Psychology*, *159*, 296–309. doi:10.1016/j.jecp.2017.02.011
Serpell, Z. N., & Esposito, A. G. (2016). Development of executive functions: Implications for educational policy and practice. *Policy Insights from the Behavioral and Brain Sciences*, *3*(2), 203–210. doi:10.1177/2372732216654718
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference/William R. Shedish, Thomas D. Cook, Donald T. Campbell*. Boston: Houghton Mifflin.
Simon, J. R., & Wolf, J. D. (1963). Choice reaction times as a function of angular stimulus-response correspondence and age. *Ergonomics*, *6*(1), 99–105. doi:10.1080/00140136308930679
Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In S. Leinhart (Ed.), *Sociological Methodology* (pp.290–312). San Francisco: Jossey-Bass.
Steele, J. L., Slater, R. O., Zamarro, G., Miller, T., Li, J., Burkhauser, S., & Bacon, M. (2017). Effects of dual-language immersion programs on student achievement: Evidence from lottery data. *American Educational Research Journal*, *54*(1_suppl), 282S–306S. doi:10.3102/0002831216634463
Valian, V. (2015). Bilingualism and cognition. *Bilingualism: Language and Cognition*, *18*(1), 3–24. doi:10.1017/S1366728914000522
Watzinger-Tharp, J., Swenson, K., & Mayne, Z. (2018). Academic achievement of students in dual language immersion. *International Journal of Bilingual Education and Bilingualism*, *21*(8), 913–928.
Weschler, D. (1974). *Weschler intelligence scale for children-revised*. New York, NY: Psychological Corporation.
Weschler, D. (1999). *Weschler Abbreviated Scale of Intelligence (WASI)*. San Antonio: Psychological Corporation.
Yang, H., Hartanto, A., & Yang, S. (2016). The importance of bilingual experience in assessing bilingual advantages in executive functions. *Cortex*, *75*, 237.