The missing explanation of the false-belief advantage in bilingual children: a longitudinal study

Vanessa Diaz1 | M. Jeffrey Farrar2

1Department of Psychology, Virginia Polytechnic and State University, Blacksburg, Virginia, USA
2Department of Psychology, University of Florida, Gainesville, Florida, USA

Correspondence
M. Jeffrey Farrar, Department of Psychology, University of Florida, Gainesville, Florida 32611, USA. Email: farrar@ufl.edu
or Vanessa Diaz, Psychology Department, Virginia Tech, Blacksburg, VA 24061, USA Email: vdiaz@vt.edu

Abstract
Bilingual preschoolers often perform better than monolingual children on false-belief understanding. It has been hypothesized that this is due to their enhanced executive function skills, although this relationship has rarely been tested or supported. The current longitudinal study tested whether metalinguistic awareness was responsible for this advantage. Further, we examined the contributions of both executive functioning and language ability to false-belief understanding by including multiple measures of both. Seventy-eight children (n = 40 Spanish-English bilingual; age M = 49.29, SD = 7.38 and, n = 38 English monolingual; age M = 47.75, SD = 6.86) were tested. A year later the children were tested again (n = 22 bilingual, n = 25 monolingual). The results indicated that language and executive function (inhibitory control) at time 1 were related to false belief in monolinguals at time 2. In contrast, bilinguals’ metalinguistic performance at time 1 was the sole predictor of false belief at time 2. The different linguistic and cognitive profiles of monolinguals and bilinguals may create different pathways for their development of false-belief understanding. A video abstract of this article can be viewed at: https://youtu.be/vILn2gKjFxw

RESEARCH HIGHLIGHTS
- Bilingual preschoolers have been reported to have an advantage in false-belief reasoning which has been attributed to their executive functioning advantages.
- The limited research exploring this relationship has not supported this connection.
- We examined the role of metalinguistic awareness, as well as language and executive functioning, in false-belief reasoning in a longitudinal study of bilingual and monolingual preschoolers.
- Metalinguistic awareness uniquely predicted later false-belief reasoning in bilinguals, whereas, language and executive function were related to monolinguals’ reasoning.

1 | INTRODUCTION

The development of false-belief (FB) understanding is a significant theory of mind achievement during the preschool years. Considerable research has demonstrated that language development and executive function contribute to its development in monolingual children (Devine & Hughes, 2014; Milligan, Astington, & Dack, 2007). Recently, several studies have reported that bilingual children pass FB tasks earlier than monolinguals (Chan, 2004; Goetz, 2003; Kovács, 2009). This bilingual advantage in FB understanding has been attributed to their enhanced executive function abilities (Goetz, 2003; Kovács, 2009), particularly inhibitory control, compared to monolinguals. Notably, only one published study has examined this explanation and did not support it (Nguyen & Astington, 2014). Similarly, while language is linked to FB understanding in monolinguals, this connection has not been examined in bilinguals (see Gordon, 2015, for an exception).

Thus, the determinants of bilingual children’s advantage in FB performance remain unidentified. The current longitudinal study of Spanish-English preschoolers explores whether metalinguistic awareness is responsible for their FB advantage. Bilingual children typically do better than monolinguals on a variety of metalinguistic awareness tests since their exposure to different languages facilitates the recognition of different ways of communicating. This may in turn lead to
understanding of variations in mental states, such as FBs. We also further explored the role of executive functioning and language ability by including multiple measures of both for a comprehensive assessment.

1.1 | Cognitive and linguistic predictors of false-belief understanding

1.1.1 | False-belief

As noted, bilinguals outperform monolingual children in understanding FBs. Goetz (2003) compared Chinese-English bilinguals to both English and Chinese monolinguals on FB reasoning. Bilinguals outperformed both groups of monolinguals, and the two monolingual groups' performance did not differ from one another in spite of coming from two vastly different language and cultural backgrounds. In a similar fashion, Kovács (2009) found that 3-year-old Hungarian-Romanian bilinguals answered FB reasoning tasks correctly more often than Romanian monolinguals. Similar, advantages in FB reasoning have also been found in French-English bilinguals (Nygren & Astington, 2014). In most of these studies, this bilingual advantage has been hypothesized as due to their enhanced executive function abilities.

1.1.2 | Executive functioning

Executive function reflects several basic processes such as inhibitory control, cognitive flexibility, planning, categorization, working memory, and following directions among others (Zelazo, Müller, Frye, & Marcovitch, 2003). Of these processes, inhibitory control and cognitive flexibility are believed to be the main executive function components involved in FB reasoning. Specifically, cognitive flexibility is thought to be involved in the child's ability to hold dual representations of the location of the object; its true location, and the mistaken belief. In turn, inhibitory control is required in FB reasoning because the child must then inhibit their own correct knowledge to correctly assess the FBs of others (see Devine & Hughes, 2014). Importantly, most of this research has focused on monolingual speakers. Carlson and Moses (2001), for example, found that inhibitory control and FB reasoning were strongly correlated even after controlling for age and language ability. In their study, inhibitory control tasks were either conflict tasks (requiring the children to inhibit a preponderant response in favor of a conflicting novel one) or delay tasks (requiring the children to delay a preponderant response, such as not peeking while the experimenter wrapped a gift). The conflict tasks were more strongly correlated with FB tasks than the delay tasks. A recent meta-analysis on the relation between FB reasoning and executive function concluded that across studies, there is a moderate but significant association between executive function and FB tasks (Devine & Hughes, 2014). For example, Hala, Hug, and Henderson (2003) found that the Dimensional Change Card Sort (DCCS) task and Luria’s tapping tasks which have both inhibitory control and working memory demands (sometimes also categorized as cognitive flexibility) were the most predictive of performance on the FB tasks. In contrast, EF tasks that only had working memory demands or inhibitory control demands were not consistently related to FB performance. Of these two, working memory was linked to FB to a greater degree (see also e.g., Cheung, 2006; Slade & Ruffman, 2005).

Bilingual children often perform better than monolingual children on certain executive function tasks (e.g., Carlson & Meltzoff, 2008). In one study, 8-year-old Canadian monolinguals were compared to Canadian bilinguals and Indian bilinguals (Bialystok & Viswanathan, 2009). Monolinguals outperformed bilinguals on English language proficiency tests and there were no differences between the two bilingual groups and the monolingual group on short-term memory and working memory tasks. In contrast, on executive functioning tasks assessing response suppression, inhibitory control, and switching; bilingual children outperformed monolinguals on tasks requiring inhibitory control and switching. There were no differences on response suppression tasks.

In a study of younger children, 5-year-old English monolinguals were compared to Spanish-English bilinguals and English speakers enrolled in a second-language immersion program (Carlson & Meltzoff, 2008). After controlling for differences in expressive vocabulary and socioeconomic status between the groups, Spanish-English bilinguals outperformed English monolinguals and the language immersion group on inhibitory control tasks such as Simon says, and on cognitive flexibility tasks, such as the DCCS task. The latter involves switching rules for sorting cards according to different dimensions like shape and color. There were no differences between the groups on the impulse-control delayed gratification tasks. An executive function advantage has also been reported in 7-month-old bilingual infants, who switched the direction of their predictive gaze faster when a stimulus began to appear from a different location, compared to monolingual infants (Kovács & Mehler, 2008). Not all studies, however, have found a bilingualism advantage in executive function tasks in children (Nygren & Astington, 2014) or adults (Hilchey, Klein, 2011; Kousaie & Phillips, 2012a, 2012b). Other studies have only found a difference between bilinguals and monolinguals after controlling for verbal ability (Carlson & Moses, 2008; Nygren & Astington, 2014). As suggested by many previous studies, vocabulary is related to most executive function tasks to some degree, but each makes independent contributions to FB understanding (Carlson & Meltzoff, 2008; Devine & Hughes, 2014).

Drawing together the relationship between executive functioning and FB reasoning in monolingual children, as well as bilinguals’ advantage on both tasks; many researchers have attributed bilinguals’ advantage in FB to their enhanced executive function (e.g., Goetz, 2003; Kovács, 2009). However, in the only published study to address this mechanism of action, this was not the case (Nygren & Astington, 2014). In this study, French-English bilingual children outperformed monolinguals in the FB tasks, but not in the executive functioning tasks. These tasks included the Day/Night Stroop conflict inhibition task, and the Backward Word span task assessing working memory. As described by the authors, performance on both these tasks has been correlated with FB reasoning in previous studies. Further, performance on the executive function was not related to the FB understanding for bilingual children. These findings indicate that an explanation for
bilinguals' advantage on FB reasoning tasks has not been identified and remains largely unexplored.

1.1.3 | Language

Language is also linked to FB understanding for monolinguals, although which specific language component (if any) is responsible has been debated. De Villiers argued that complementation syntax provides the cognitive structures and tools needed to represent and reason about FB (de Villiers, 2005). In this grammatical structure, a sentence that contains a mental (think) or communication verb (say) is followed by an embedded clause or complement that allows a distinction between reality and the belief of the speaker. For example, "Sally thinks that the marble is in the basket, but it really is in the box" is an example of a complementation sentence in that although the reality of the first clause is false, the sentence as a whole can be true. Memory for complementation syntax is correlated with FB reasoning capacities. In a meta-analysis by Milligan et al. (2007), the largest effect size in predicting FB task performance was due to memory for complements, although all language measures were significant contributors, including vocabulary and general syntax. In their analysis, however, only a small number of studies investigated the relationship between FB reasoning and memory for complementation. Moreover, the effect has not been reliably replicated (Cheung, Hsuan-Chih, Creed, Ng, Wang, & Mo, 2004; Farrar, Lee, Cho, Tamargo, & Seung, 2013).

Other approaches have argued that general language (semantic and/or syntactic abilities) is sufficient for FB understanding, and complementation is not required (Slade & Ruffman, 2005). One proposal is that it is through linguistically mediated social interactions that children discover that other people have different minds (Carpendale & Lewis, 2004; Fernyhough, 2008; Nelson, 2005, 2010). For example, maternal mental state talk is related to FB task performance (Ruffman, Slade, & Crowe, 2002).

Language ability has been a contentious topic in which to compare bilinguals and monolinguals. Bilingual children are faced with two competing linguistic groups creating phonetic, semantic, and grammatical conflicts. The bilingual child, for instance, needs to recognize that the same object can have two phonetically distinct labels in order to acquire semantic knowledge. These conflict demands are not experienced nearly as often by monolinguals. It is not surprising then that bilingual children often score lower on language ability tests compared to monolinguals (e.g., Bialystok & Viswanathan, 2009; Carlson & Meltzoff, 2008), although combining bilinguals’ vocabulary scores on measures of both of their languages reduces performance differences between the groups (Core, Hoff, Rumiche, & Señor, 2013). Studies of language proficiency and its relation to FB reasoning in bilingual children are limited. In one study, vocabulary proficiency in both Spanish and English was associated with bilingual children’s FB performance (Gordon, 2015). Whether complementation or grammar is related to FB reasoning in bilingual children remains unknown. Thus, there is limited evidence that executive function, particularly inhibitory control, or language ability is responsible for bilingual children's FB advantage, and will be examined in the current study.

1.1.4 | Metalinguistic awareness

A third possibility is that metalinguistic awareness accounts for bilingual children’s FB performance (Goetz, 2003). Metalinguistic awareness reflects knowledge of the formal properties of language as an object of thought. Different aspects of metalinguistic awareness have been measured in preschool and school-aged children, including phonological awareness (Farrar & Ashwell, 2012), definitional abilities (Bialystok, 1988), understanding of label referent relations (Ben-Zeev, 1977), ratings of grammaticality (Bialystok & Barac, 2012), and understanding of homonyms and synonyms (Doherty, 2000; Doherty & Perner, 1998; Graham, Brooks, Graham, & Ostenfeld, 2000). Metalinguistic awareness reflects a more flexible understanding of language, including the capacity to hold conflicting representations, and thus many tasks require both inhibitory control and cognitive flexibility.

Studies of monolingual children have suggested that FB tasks reflect cognitive flexibility in understanding meta-representations, allowing for the emergence of metalinguistic awareness. In a homonym task, for example, children must be able to flexibly shift between two meanings of the same phonological word, such as bat. Thus for monolingual children, FB provides the cognitive flexibility that can be applied to metalinguistic tasks (Doherty & Perner, 1998). While there are no longitudinal studies of the relationship between FB understanding and metalinguistic performance in monolingual children, the conceptual argument is that FB provides the cognitive foundation for metalinguistic awareness (Doherty, 2000).

Similar to findings of executive function, bilinguals often outperform monolinguals on metalinguistic awareness tasks (Bialystok, Majumder, & Martin, 2003). In one of the earliest accounts, Ben-Zeev (1977) reported a bilingual advantage when children were asked to switch familiar labels and to answer questions using the new labels (such as between sun and moon). Similar bilingualism effects were found using the same sun/moon task, as well as a label/referent independence task where children were asked, for example, if one can describe a giraffe, even if all giraffes in the world were to die (Cummins, 1978). Chan (2004) also reported bilingualism effects using a moving label task, where pre-literate children were asked what a label read after it was moved from under one illustration to another.

Of interest in the present study is how metalinguistic awareness contributes to the emergence of FB understanding in bilingual children compared to monolingual children. There is hardly more substantial evidence that the world is filled with agents with active mental worlds than that of language, which is used precisely to express conscious mental activity. Language provides evidence of “minds in action” (Schick, de Villiers, de Villiers, & Hoffemeister, 2007). Bilingual children may first discover differences in mental activity as they experience words, meaning, and structure in the two languages they are acquiring, thus leading to the metalinguistic awareness that speakers of different languages manifestly know different things. Without the proper representational abilities, children might interpret language as a direct portrayal of the world, rather than as evidence of cognitive representations on the part of the speaker. Once the arbitrariness
and representational nature of language becomes apparent, children can then begin to see it as a tool that can be used to represent and interpret mental states. Thus, in contrast to the monolingual argument that false belief predicts metalinguistic awareness, for bilingual children, it is hypothesized that metalinguistic awareness may predict FB development.

In summary, bilingual children outperform monolinguals on FB tasks (Goetz, 2003; Kovács, 2009; Nguyen & Astington, 2014), on executive functioning tasks (Bialystok & Viswanathan, 2009; Carlson & Meltzoff, 2008; Kovács & Mehler, 2008), and on metalinguistic awareness tasks (Ben-Zeev, 1977; Chan, 2004; Cummins, 1978). It is also the case that executive function, language ability, and metalinguistic awareness are related to FB reasoning in monolingual children, although the direction of influence may vary depending on the cognitive measure. In the present longitudinal study, we explore the influence of these factors on the emergence of FB understanding in both bilingual and monolingual preschool children. Of interest is whether the different cognitive profiles of monolingual and bilingual children lead to different pathways to FB understanding in these two groups. In fact, studies on different populations, such as children with autism and those who are deaf, have identified different linguistic routes to FB understanding (Farrar, Benigno, Tompkins, & Gage, 2017).

Two research questions were of interest. First, were there performance differences between bilinguals and monolinguals on executive function, false-belief, metalinguistic awareness and language measures? It was expected that monolinguals would outperform bilinguals on language proficiency at both time points, whereas bilinguals would outperform monolinguals on FB reasoning, executive function, and metalinguistic awareness.

Second, how do language, executive function, and metalinguistic awareness predict later FB reasoning for bilinguals and monolinguals? We were interested in these relationships at each of the two time points, but primarily in the time 1 predictors of FB reasoning at time 2. For bilinguals, time 1 metalinguistic awareness was expected to be the strongest predictor of FB reasoning at time 2, followed by executive function. For monolinguals, language ability and executive functioning at time 1 were hypothesized to be related to FB reasoning at time 2.

2 | METHOD

2.1 | Participants

Seventy-eight children (n = 40 Spanish–English bilingual, n = 38 English monolingual) aged 35 to 66.80 months (bilingual M = 49.29, SD = 7.38, monolingual M = 47.75, SD = 6.86) were recruited from preschools in predominantly bilingual and predominantly monolingual communities in the United States. Bilingual children were identified by their parents as being fluent in both Spanish and English, and regularly interacting with speakers of both these languages. The vast majority of the bilingual children (77.5%) had been exposed to both languages since birth, and all had been exposed to their non-dominant language for at least one year (only three had only one year of immersion in English). Parents also reported that in 87.5% of bilingual children’s homes either both or a language other than English were spoken, while 67% reported that in the preschool the child attends both languages were spoken regularly. In addition, 59% of parents reported reading to their children in both languages. In turn, 26% of mothers of bilingual children speak to them in both languages, as do 31% of fathers. Analyses indicated no differences between bilinguals and monolinguals on socioeconomic status (SES) as reflected in maternal education, occupation, and income, using a series of independent samples t tests, ps > .05.

Available children were tested again a year later (n = 22 Spanish–English bilingual [M = 56.56, SD = 5.22], n = 25 English monolingual [M = 56.85, SD = 6.58]). Attrition was due to children either moving away or changing schools. Testing was conducted during the summer sessions, resulting in less consistency in enrollment and attendance compared to the formal school year. There were no differences between the children who were tested at only the first time point compared to those tested at both sessions on time 1 receptive vocabulary, false-belief understanding, metalinguistic awareness, and executive functioning, all ps > .05. The only exception was that children who were not tested at time 2 were slightly older (M = 51.4 months) at time 1 than children assessed at both time periods (M = 47.5 months) (F(1, 45) = 10.66, p < .05. Time 1 age was controlled in subsequent analyses.

2.2 | Measures

2.2.1 | Theory of mind

Children were administered an unexpected location task (Wimmer & Perner, 1983), unexpected content task, and an appearance–reality task (involving a sponge-rock) using standard testing protocols. Questions from each of the tasks were classified into whether they were questions about FBs in the other (False-Belief composite 0–4), FBs in the self (Representational-Change composite 0–2), and awareness of the true identity of an object (Appearance–Reality composite 0–1). These composites were then added to form an overall FB reasoning measure. Memory control questions were asked for each task and were required to receive credit for correct answers. A total FB score was calculated (range 0–7).

2.2.2 | Executive functioning

A series of well-established inhibitory control and cognitive flexibility measures were administered according to standard format. Measures were selected based on prior studies showing bilingual and monolingual group differences on conflict tasks (Carlson & Meltzoff, 2008), as well as their relation to FB reasoning tasks (Devine & Hughes, 2014).

Bear/dragon Simon says-like task (Reed, Pien, & Rothbart, 1984) is an inhibitory control measure. The child received one point for ignoring each of the dragon commands, and one point for following each of the bear commands for a total of 10 points.

Happy-Sad Stroop-like task (Lagattuta, Sayfan, & Monsour, 2011, based on Gerstadt, Hong, & Diamond, 1994, Day/Night task) is an inhibitory control measure asking children to say “happy” when a “sad”
face is shown, and vice versa. The final score was the number of correct responses ranging from 0 to 16 points.

Dimensional Change Card Sort task (DCCS) (Frye, Zelazo, & Palfai, 1995) is a measure of cognitive flexibility. The child's score was computed out of 4 possible points based on the correct sorting of the red boat (twice) and blue rabbit (twice) in the second set, after the sorting dimension was changed from color to shape. The other cards (blue boat and red rabbit) were congruent with the model on both dimensions.

Composite: An executive function composite score was created by first transforming each of the three task scores into z-scores due to the differing number of trials per task. The composite was formed by adding the individual z-scores.

2.2.3 Memory

Short-term memory-digit span (Bialystok, 2010) task measures children's short-term memory span. The experimenter read single digit lists at a rate of one digit per second starting with two digits. The child was asked to repeat the numbers in the same order. One digit was added after each second trial until the child was unable to reproduce both trials of the same number of digits. The final score consisted of the last number of digits in which the child was able to reproduce at least one sequence.

2.2.4 Metalinguistic awareness

Semantic Rhyming Task (adapted from Farrar & Ashwell, 2012). For the semantic rhyming task the child was presented with six sets of three pictures each. Two of the pictures rhymed while the third did not. The third picture, however, had a semantic association to one of the other two pictures (e.g., truck-car-duck). The child was asked to point to the two words that rhyme. Three practice trials were given with help from the experimenter by giving the child the correct answer and explanation. There were six trials in this task and scores ranged from 0 to 6.

Symbol Substitution Test (adapted from Ben-Zeev, 1977). This task required the child to recognize that a word can be substituted for another: “You know that in English this is named airplane [experimenter showed a picture of an airplane]. In this game its name is apple... Can the apple fly? [Correct answer: Yes.] How does the apple fly? [Correct answer: With its wings.]” Five word pairs were used for a total of five trials with two questions in each trial for a total of 10 points.

Synonym Judgment Task (Doherty & Perner, 1998). Children were first asked to identify a set of nine cards by pointing to the correct picture card called out by the experimenter (E). Each card was called out twice using two different labels (i.e., couch and sofa), in alternating order. Three practice trials followed in which the child was asked to name one of the pictures, and then judge a puppet's response when the puppet was asked to “say the other name”. For example, the child was shown a picture of a ship and told, “This can be called either a ship or a boat. What would you like to call it?” After the child's response, E brought out a puppet and said to the puppet, “Now, Puppet, you say the other name”. The puppet then used the same name as the child, for example, “ship”, and E then asked, “Is that what he should have said?” If correct, E continued: “No, because you said ‘ship’ didn’t you? Very good!” Four trials were given with no feedback after the three practice trials. In two of these trials the puppet used the synonym correctly, in another the puppet used the same word as the child, and in another the puppet used an incorrect label. The child was given 1 point each time she judged the puppet's responses correctly for a total of 4 points. Some of the synonyms were hyponyms (i.e., rose and flower), as it proved difficult to find pure synonyms for physical objects familiar to 3-year olds.

Homonym Selection Task (Doherty, 2000). A vocabulary check was first performed to make sure children were familiar with both uses of the four target words. In order to do this, children were asked to identify each of the four target homonyms from a set of four picture arrangements each; with each target word appearing twice depicting a different object each time (eight sets in total plus four more for the two practice words). For example, one set of pictures contained a picture of a baseball bat, and another contained a picture of a flying mammal bat. In both instances the children were asked, “Can you point to the bat?” The practice and test trials then proceeded with both pictures of each word appearing alongside two other pictures in similar four-card arrangements (for example, a baseball bat, a cake, a flying mammal bat, and a shirt). E then pointed to and named an extra card depicting one of the two homonym pictures in the arrangement (e.g., the baseball bat). The child was then asked to point to “the other bat”. After two practice trials, four trials using different homonyms were given, for a total of 4 points.

Composite. A metalinguistic awareness composite score was created by transforming each task score into z-scores due to the differing number of questions per task. The composite was created by adding the individual z-scores.

2.2.5 Language

Expressive One Word Picture Vocabulary Test (EOWPVT). Children's expressive vocabulary was measured using the English monolingual and Spanish-English bilingual versions of the EOWPVT (Bronwell, 2000). Raw scores were used in the analyses.

Receptive One Word Picture Vocabulary Test (ROWPVT). Children's receptive vocabulary was measured using the English monolingual and Spanish-English bilingual versions of the ROWPVT. Raw scores were used in the analyses (Gardner, 1985).

Complementation Task (de Villiers & Pyers, 2002). Comprehension for complementation was measured in the following manner: E read to the child six sentences with embedded complements, accompanied with illustrative pictures. E then asked a comprehension question such as: “She thought the girl was reading a book, but she was really playing cards”; the comprehension question would be, “What did she think?” Although the standard task used 12 items, we used a smaller subset to reduce testing time. The child's score was a total of 6 possible points. Half the sentences contained a mental state verb (think) and half contained a communication verb (say). Complementation functions similarly in English and Spanish.
2.3 | Procedure

2.3.1 | Time 1

Parents provided informed consent through signed letters in English for the English monolingual parents, and in either Spanish or English for parents of the bilingual children. Parents also completed a demographic questionnaire to help determine their SES and possible developmental delays in the child. None of the latter were reported. In addition, they were asked to fill out a language questionnaire that asked about the child’s exposure to different languages in different contexts such as home and school, and their preferred language. All families were compensated with a $10 gift card for each of the two sessions for a total of $20. The testing was conducted in play-like interactions in quiet rooms at the children’s preschools such as offices and classrooms not being used by others while the testing sessions were ongoing. Each session lasted approximately 40 minutes. No sessions had to be discontinued because of fatigue or non-cooperation from the children.

All measures were given in English to English monolinguals, and to bilinguals in the language selected in the parent questionnaire as the child’s dominant language. The experimenter corroborated bilinguals’ language dominance using an expressive vocabulary measure during the first testing session. Over a third (37%) of bilingual children were tested in Spanish. All vocabulary measures were given to bilingual children first in their non-dominant language and then in their dominant language. The testing sessions began with a language test that lasted approximately 10 minutes. Testing then proceeded with FB, executive functioning, and metalinguistic awareness tasks in alternating order. The tasks were given to all children in a fixed sequence as per convention in the literature.

2.3.2 | Time 2

A year after time 1 data were collected, the children were tested again. The same language tests were used, as were the executive functioning and metalinguistic awareness tasks. In turn, the FB tasks were changed slightly to preserve the element of surprise as appropriate.

3 | RESULTS

Preliminary analyses indicated no differences between bilinguals and monolinguals on demographic variables of age, gender, maternal education or income, at either time point using a series of independent samples t tests.

3.1 | Monolingual and bilingual performance differences

3.1.1 | Time 1

We first examined whether there were performance differences between bilinguals and monolinguals. Means and standard deviations of the FB and language measures are presented in Table 1. As expected, at time 1 monolingual children outperformed bilinguals in language ability in both receptive and expressive vocabulary (ROWPVT t(76) = 5.49, p < .001; EOWPVT t(76) = 6.90, p < .001). It is important to note that these vocabulary measures for the bilingual children reflect their combined scores for knowing a word in either English or Spanish (Core et al., 2013). On the complementation measure, both groups scored equivalently. For FB understanding, there were no group differences. However, once the significant difference in receptive vocabulary was controlled, bilinguals demonstrated an advantage in FB reasoning abilities (F(1, 77) = 8.39, p < .01, \( \eta^2 = .106 \)) with estimated means provided in Table 1.

For executive functioning, there were no differences between monolinguals and bilinguals in the raw scores or the composite measure as shown in Table 2. As with FB, once vocabulary was controlled, bilinguals performed better than monolinguals on the composite
executive functioning measure ($F(1, 76) = 11.81, p < .01, \eta^2 = .183$). Similar group differences were found for the DCCS and bear dragon tasks (all $F$s(1, 76) > 4.45, $p < .05, \eta^2 > .06$). The happy/sad inhibitory control measure and short-term memory did not differ between the groups even when controlling for vocabulary.

As indicated in Table 3, bilingual children outperformed monolinguals on the composite metalinguistic awareness measure ($F(1, 76) = 16.63, p < .001, \eta^2 = .18$). There were no differences on raw individual metalinguistic measures, except for the symbol substitution task with bilinguals performing higher ($t(76) = -4.08, p < .001$). Once receptive vocabulary was controlled, bilinguals outperformed monolinguals on the synonym, rhyming, and homonym tasks, all $F$s(1, 76) > 6.0, $p < .01, \eta^2 > .18$. The estimated means are provided in Table 3.

### 3.1.2 | Time 2

A year later, children were tested again ($n = 22$ Spanish-English bilingual [M = 56.56, SD = 5.22], $n = 25$ English monolingual [M = 56.85, SD = 6.58]), with a similar proportion of males and females. Means and standard deviation of the language and FB measures are presented in Table 1. Mean comparisons revealed that, again, monolingual children outperformed bilingual children on vocabulary measures (ROWPVT $t(45) = 4.50, p < .001$; EOWPVT $t(45) = 6.64, p < .001$), but not in complementation.

No differences were detected between the groups on FB or on the executive functioning composite or individual measures, except for short-term memory ($t(45) = 2.96, p < .01$) with monolinguals outperforming bilinguals (see Table 2). Once controlling for vocabulary, bilinguals outperformed monolinguals on the metalinguistic awareness composite ($F(1, 45) = 8.38, p < .01, \eta^2 = .176$). There were no differences in the individual metalinguistic measures at time 2 (see Table 3) and the results did not change when controlling for differences in vocabulary or short-term memory.

### 3.2 | Correlates and predictors of FB

The main research question examined the relationships between language, executive function, and metalinguistic awareness to FB reasoning for bilinguals and monolinguals. Of primary importance were the predictors of time 2 FB reasoning from the time 1 language, executive function and metalinguistic measures. The relationships between the variables within each time point were explored as well.

All of the reported time 1 partial correlations controlled for age and are presented in Table 4. The composite measures for executive function and metalinguistic awareness, as well as the individual tests were used. At time 1 for monolinguals, FB reasoning was correlated with executive function (Bear Dragon task) and the metalinguistic awareness composite. Controlling for the Bear Dragon task, the FB and metalinguistic comparison remained significantly correlated ($r(32) = .36, p < .05$). For bilinguals, FB was significantly associated with metalinguistic awareness and expressive vocabulary. This relationship was maintained even after controlling for vocabulary ($r(32) = .38$).

Similar partial correlations controlling for age were conducted at time 2. For FB reasoning, a very distinct pattern emerged for the language groups. As shown in Table 5, monolinguals’ FB time 2 was significantly related to the time 2 executive functioning composite and language ability (receptive language and complementation). For bilinguals, time 2 FB was only related to time 2 metalinguistic awareness.

Of primary interest were the time 1 predictors of time 2 FB performance for bilingual and monolingual preschoolers. For monolinguals, time 1 executive function (card sort $r(21) = .42, p < .05$), complementation ($r(21) = .52, p < .05$), and receptive vocabulary ($r(20) = .40, p =$
.052) were significant predictors of FB at time 2, as shown in Table 6. For bilinguals, the time 1 metalinguistic awareness composite was predictive of FB reasoning at time 2 (r(18) = .49, p < .05), although the effect was driven mainly by the synonym judgment task at r(18) = .72, p < .001. None of the other time 1 measures predicted later FB reasoning in the bilingual population.

Finally, we examined the bidirectional relation of FB and metalinguistic awareness for both populations. As shown in Figure 1, for monolinguals time 1 FB predicted time 2 metalinguistic awareness, as it did bilinguals. The time 1 predictors of FB at time 2 differed between groups. Only metalinguistic awareness made a unique contribution to bilinguals’ time 2 FB performance, whereas, for monolingual preschoolers metalinguistic awareness at time 1 did not make a unique contribution once language and executive function were controlled. Thus, the relationship between FB and metalinguistic awareness from time 1 to time 2 was bidirectional for bilingual children, but unidirectional for monolingual preschoolers.

### Discussion

In this study we examined the emergence of FB reasoning in bilingual preschoolers. Of particular interest was identifying the predictors for the reported FB advantage in bilingual children. A number of factors have been associated with FB reasoning in monolingual preschoolers, such as language, executive functioning and metalinguistic awareness. Further, monolinguals and bilinguals often differ in their performance on these measures. In the current longitudinal study, we examined

### Table 3

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Synonym</th>
<th>Homonym</th>
<th>Rhyming</th>
<th>Symbol substitution</th>
<th>Metalinguistic composite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilinguals (n = 38)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>3.07 (1.03)</td>
<td>3.44 (1.05)</td>
<td>3.10 (1.83)</td>
<td>6.39 (3.34)*</td>
<td>.80 (2.84)*</td>
</tr>
<tr>
<td>Estimated M*</td>
<td>3.28 (.16)*</td>
<td>3.83 (.19)*</td>
<td>3.47 (.29)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monolinguals (n = 35)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>2.92 (.92)</td>
<td>2.97 (1.48)</td>
<td>2.78 (1.73)</td>
<td>3.11 (.65)*</td>
<td>-.77 (2.62)*</td>
</tr>
<tr>
<td>Estimated M*</td>
<td>2.68 (.17)*</td>
<td>2.54 (.20)*</td>
<td>2.37 (.31)*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time 2</th>
<th>Synonym</th>
<th>Homonym</th>
<th>Rhyming</th>
<th>Symbol substitution</th>
<th>Metalinguistic composite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilinguals (n = 22)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>3.60 (.76)</td>
<td>3.76 (.60)</td>
<td>4.04 (1.74)</td>
<td>6.16 (3.75)</td>
<td>.15 (2.63)*</td>
</tr>
<tr>
<td>Estimated M*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monolinguals (n = 25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>3.70 (0.96)</td>
<td>3.63 (0.88)</td>
<td>4.00 (1.54)</td>
<td>4.41 (3.90)</td>
<td>-.20 (2.7)*</td>
</tr>
<tr>
<td>Estimated M*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Estimated means and standard errors obtained from ANCOVA controlling for language ability.

Z-transformed score means, standard deviations, estimated means, and standard errors.

*p < .05; significant difference between bilinguals and monolinguals at each time point.

### Table 4

<table>
<thead>
<tr>
<th>Time 1</th>
<th>FB</th>
<th>EF</th>
<th>Express</th>
<th>Recep</th>
<th>Compl</th>
<th>Metalinguistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilinguals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FB1</td>
<td>—</td>
<td>.21 (.38)*</td>
<td>.27</td>
<td>.23</td>
<td>.20</td>
<td>.42^</td>
</tr>
<tr>
<td>EF1</td>
<td>.21</td>
<td>—</td>
<td>.13</td>
<td>.41*</td>
<td>.25</td>
<td>.46*</td>
</tr>
<tr>
<td>Expressive1</td>
<td>.41*</td>
<td>.30^</td>
<td>—</td>
<td>.38*</td>
<td>.32^</td>
<td>45^ (.37)*</td>
</tr>
<tr>
<td>Receptive1</td>
<td>.27</td>
<td>.15</td>
<td>.63*</td>
<td>—</td>
<td>.07</td>
<td>.25 (.40*^)</td>
</tr>
<tr>
<td>Complement1</td>
<td>.30^</td>
<td>.21</td>
<td>.28^</td>
<td>.30^</td>
<td>—</td>
<td>.22</td>
</tr>
<tr>
<td>Metalinguistic1</td>
<td>.34*</td>
<td>.11</td>
<td>.19 (.37)*</td>
<td>.22 (.40)*</td>
<td>.05</td>
<td>—</td>
</tr>
</tbody>
</table>

| Monolinguals | | | | | | |

EF = Executive Functioning, Express = Expressive Vocabulary, Recept = Receptive Vocabulary; Compl = complementation.

Monolinguals upper diagonal.

Bilinguals lower diagonal.

^Bear Dragon task.

^SYNonym task.

^p < .10; *p < .05; **p < .001.
whether the predictors of FB bilinguals’ metalinguistic awareness accounted for their FB performance, which differed from monolinguals’ FB development. As far as we know, this is the first longitudinal study of FB development in bilingual children.

### 4.1 Bilingual–monolingual performance differences

As reported previously with other bilingual populations, at time 1 the Spanish-English bilinguals in the current study had an advantage in FB reasoning after controlling for receptive vocabulary. This result generally confirmed the pattern of previous research (Goetz, 2003; Kovács, 2009; Nyguen & Astington, 2014). Also consistent with prior research, bilingual children had smaller expressive and receptive vocabulary scores than monolingual children. At the same time, however, bilinguals significantly outperformed monolinguals on the metalinguistic awareness composite. These results corroborate the previously reported advantage that bilinguals exhibit in metalinguistic awareness (see Ben-Zeev, 1977; Cummins, 1978; Hermanto, Moreno, & Bialystok, 2012).

To further explore bilingual–monolingual performance differences, we controlled statistically for the significant language imbalance between the groups, and found that bilinguals outperformed monolinguals on the FB reasoning composite (see also Nyguen & Astington, 2014), as well as on the executive function composite (see Bialystok & Viswanathan, 2009; Carlson & Meltzoff, 2008; Kovács & Mehler, 2008; Martin-Rhee & Bialystok, 2008). Since bilinguals tend to perform lower on language measures, group differences in executive functioning or theory of mind measures are often examined by controlling for language (Carlson & Meltzoff, 2008; San Juan & Astington, 2012). This has been controversial because it could either indicate that bilinguals have better executive function and/or FB reasoning skills that are masked by their language deficits, or that the executive function and FB skills between monolinguals and bilingual groups are equivalent as indicated by their raw scores. Future studies should explore this issue by using measures that have low linguistic demands.

### Differences in vocabulary knowledge were maintained at the second time point with bilinguals underperforming compared to monolinguals on both receptive and expressive vocabulary measures. Bilinguals also underperformed on the short-term memory task but only at time 2. Controlling for short-term memory, however, did not lead to any differences in the group comparisons for FB or executive function. In contrast, controlling for language ability at time 2 led to differences between bilinguals and monolinguals on the metalinguistic awareness composite, with bilinguals outperforming monolinguals.

The reductions of bilingual–monolingual differences on FB reasoning and executive function at time 2 could be attributed to two disparate reasons. The first one is that a reduction in sample size could account for the loss of power to detect differences, as observed for example in that the differences in metalinguistic awareness were significant after controlling for language ability at this time point, but without statistical controls at the previous time point. The second possibility is that older children perform better on the tasks and bilingual–monolingual differences disappear as both groups perform more closely to ceiling. This was indeed corroborated by smaller standard deviations at time 2 for both groups on both measures.

---

### TABLE 5 Partial correlations (controlling for age) between FB, executive functioning, language and metalinguistic awareness at Time 2

<table>
<thead>
<tr>
<th>Time 2</th>
<th>FB 2</th>
<th>EF 2</th>
<th>Express 2</th>
<th>Recept 2</th>
<th>Compl 2</th>
<th>Metalinguistic 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilinguals</td>
<td>FB 2</td>
<td>.44*</td>
<td>.40^</td>
<td>.44^</td>
<td>.45^</td>
<td>.40^</td>
</tr>
<tr>
<td>Monolinguals</td>
<td>EF 2</td>
<td>.24</td>
<td>—</td>
<td>.22</td>
<td>.37^</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>Expressive 2</td>
<td>.32</td>
<td>.36</td>
<td>—</td>
<td>.59^</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td>Receptive 2</td>
<td>.33</td>
<td>.35</td>
<td>.55^</td>
<td>—</td>
<td>.34</td>
</tr>
<tr>
<td></td>
<td>Complement 2</td>
<td>.26</td>
<td>.23</td>
<td>.12</td>
<td>.29</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic 2</td>
<td>.55^</td>
<td>.44^</td>
<td>.53^</td>
<td>.22 (40^)</td>
<td>.19</td>
</tr>
</tbody>
</table>

EF = Executive Functioning, Express = Expressive Vocabulary, Recept = Receptive Vocabulary; Compl = complementation.


1 Bear Dragon task.
2 Synonym task.
\(^* p < .10; ^{\circ} p < .05; ^{**} p < .001.

---

### TABLE 6 Partial correlations (controlling for age) between Time 1 EF, language, and metalinguistic awareness, and Time 2 FB

<table>
<thead>
<tr>
<th>Time 1</th>
<th>FB 2</th>
<th>Express 2</th>
<th>Recept 2</th>
<th>Compl 2</th>
<th>Metalinguistic 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilinguals</td>
<td>(n = 22)</td>
<td>FB 2</td>
<td>.39^</td>
<td>.20</td>
<td>.29</td>
</tr>
<tr>
<td>Monolinguals</td>
<td>(n = 25)</td>
<td>Expressive 1</td>
<td>.32 (42^2)</td>
<td>.29</td>
<td>.40^</td>
</tr>
<tr>
<td></td>
<td>Receptive 1</td>
<td>.37</td>
<td>.09</td>
<td>.52^</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Complementation1</td>
<td>.49 (.72**)</td>
<td>.32</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^* p < .10; ^{\circ} p < .05; ^{**} p < .001.

EF = Executive Functioning.
1 Synonym task.
2 Card sort task.
One way to conceptualize the bilingual advantage on FB reasoning and executive functioning controlling for children's vocabulary is as follows: taken at face value, the reduced language proficiency we see in bilingual children should suggest an accompanying delay in FB and executive function (Nguyen & Astington, 2014). However, we did not find this delay. We suggest that the increased social complexity and metalinguistic requirements of navigating two language environments for bilingual children may strengthen their cognitive and socio-cognitive abilities. While it is controversial, statistically controlling for language ability examines whether FB and executive function performance differs if their language skills were equal (Nguyen & Astington, 2014).

4.2 Differences in the correlates and predictor of FB reasoning

Of primary interest were the predictors of FB reasoning for bilinguals and monolinguals, and whether there were differences between the groups in the pattern of these relationships. At time 1, for monolinguals the FB composite was related to the executive function Bear/Dragon task and to the metalinguistic awareness composite. For bilinguals, in turn, the FB composite was related to the metalinguistic awareness composite, and to language ability (expressive vocabulary), but not to executive function as for monolinguals. The pattern of results at time 1 suggests that the developmental pathways of FB reasoning differ for these two groups; not only in overall performance but in the correlates of FB reasoning.

These patterns were even more distinctive at time 2. For monolinguals FB reasoning was related to executive function, receptive language, and complementation. This is very much in line with the extensive literature on FB development in monolingual children (Devine & Hughes, 2014; Milligan et al., 2007). In contrast, only metalinguistic awareness correlated with bilingual preschoolers’ FB understanding at time 2.

While suggestive, same time point correlations do not indicate the direction of influence between variables. Of primary interest was whether the longitudinal relations between time 1 predictors of time 2 FB reasoning differed for the two groups. For monolinguals, executive functioning (card sorting task) and language ability at time 1 were predictive of FB reasoning a year later. The card sort task (Frye et al., 1995) is an executive functioning task that reflects cognitive flexibility. Rather than being asked to inhibit a preponderant response, children are asked to switch and sort cards along two different dimensions. In effect, this requires the ability to hold dual representations in mind, a hallmark of FB reasoning (Perner, Stummer, Sprung, & Doherty, 2002). Interestingly, time 1 FB was not related to time 2 FB for either group. One possible explanation is that there was less variability at Time 2 in FB performance.

For bilinguals, metalinguistic awareness at time 1 was the only significant predictor of FB reasoning a year later. These results were consistent with our hypothesis that metalinguistic awareness would be particularly crucial for the development of FB understanding in bilinguals. It was also predicted that executive function at time 1 would predict FB reasoning at time 2, but this was not supported. This is particularly noteworthy given previous hypotheses that it is bilinguals’ noted executive function advantage that is behind their FB performance advantage (Goetz, 2003; Kovács, 2009). This was not supported in the current study or in the Nguyen and Astington (2014) study. Even though the current study utilized several executive functioning measures, no effects emerged for executive function predicting bilinguals’ FB reasoning.
Instead, as hypothesized, metalinguistic awareness predicted bilingual children’s FB performance. There was a consistent bilingual performance advantage on the metalinguistic awareness measures in the current study. Specifically, for bilinguals, language differences between speakers are a very salient characteristic of their environment, making them more aware of the nature and characteristics of language (i.e., metalinguistic awareness). This awareness may lead in turn to the recognition of other people as producers of language engage in different mental activity that this language communicates; leading to an understanding of FB reasoning.

Finally, this relationship between FB and metalinguistic awareness was bidirectional for bilinguals. Specifically, there was a strong relationship between the entire metalinguistic awareness composite at time 1 and the FB composite at time 2 as previously discussed. However, the reverse relationship was also found; FB reasoning at time 1 significantly predicted metalinguistic awareness at time 2. For monolinguals, there was no relationship between time 1 metalinguistic awareness and time 2 FB. However, the FB reasoning composite at time 1 predicted the homonym selection task at time 2. These results suggest that for monolinguals FB reasoning at time 1 is notably more predictive of metalinguistic awareness at time 2 than the reverse. Why might this be the case? As described by Doherty and Perner (1998), the ability to handle conflicting representations about mental states such as those involved in FB tasks, may later translate into the ability to handle conflicting linguistic representations, through the ability to represent different perspectives (Doherty, 2002; Perner et al., 2002).

Thus, bilinguals’ FB reasoning is linked to their metalinguistic abilities which emerge from their experiences with two languages. Bilingual children’s performance on the metalinguistic measures reflects their varied linguistic experience. That is, through social interactions with speakers of different languages, children become aware that others have different ways of communicating using different words and different grammatical constructions. This awareness of language as a representational system that varies across speakers of different languages, leads to an understanding of FB reasoning. Moreover, the ability to handle conflicting representations about mental states such as those involved in FB tasks, may later translate into the ability to handle conflicting linguistic representations, through the ability to represent different perspectives (Doherty, 2002; Perner et al., 2002).

4.3 | Limitations and future directions

There were some limitations to this study. First, prior research has often shown that language is related to the primary variables of interest in the current study, namely FB, EF, and metalinguistic development (e.g., Astington & Baird, 2005; Carlson & Meltzoff, 2008; Devine & Hughes, 2014). Thus, the measures are not always independent of the influence of language. In the present study, we observed differential patterns of the role of language in performance depending on the population and tasks. For instance, at time 2 receptive vocabulary was related to FB and the composite executive function task for monolingual children but not bilingual children. Further, controlling for language in the ANCOVAs sometimes produced group difference on some, but not all, measures. Given the pattern of bilinguals’ linguistic development, future studies should also employ executive functioning and FB reasoning tasks with lower linguistic demands.

Second, it is important to note that not all executive function or metalinguistic awareness measures were related to FB reasoning. This may suggest that the various measures have different developmental trajectories as has been found in other studies (e.g., Carlson & Moses, 2008). In addition, we only included a short-term memory assessment. Future studies should employ a working memory task such as the Missing Scan Task (Roman, Pisoni, & Kronenberger, 2014) that may be more sensitive to working memory ability.

Third, the sample size at the second time point was smaller. Even though attrition was not systematic, some bilingual–monolingual differences that were significant at time 1 were not so at time 2. This could be due to a decrease in the differences with age, as monolinguals begin to “catch-up” to the bilingual children. A larger sample would provide more power to detect these differences. Another potential issue regarding attrition is reduced variability at time 2. For example, we observed slight reductions in the standard deviation of our measures at time 2 compared to time 1, although the scores at time 2 for the measures typically reflected the entire range.

Another possible limitation regards the classification of children as monolingual or bilingual. While the current study used a standard procedure, there are other ways to do so. For instance, bilingual children can be distinguished by whether they are sequential or simultaneous bilinguals, as well as the circumstances in which they acquire their second language. In addition, the order in which the tests were administered could have affected performance. Specifically, it is possible that having bilinguals’ expressive vocabulary tested first in both languages may have primed either their executive function or metalinguistic awareness, or in contrast, it may have exhausted them cognitively. Finally, the lack of a Spanish monolingual group for comparison may be a limitation, even though we have no particular reason to suspect that the development of FB reasoning may be different in Spanish and English.

4.4 | Conclusion

In the current study we identified metalinguistic awareness as a unique contributor to bilingual preschoolers’ FB advantage. This was in contrast to prior hypotheses that their performance was due to executive function. In contrast, for monolinguals the predictors of FB were consistent with prior research demonstrating the role of both executive function and language, particularly complementation. Further, whereas a bidirectional relationship was found between metalinguistic awareness and FB reasoning for bilinguals, for monolinguals FB reasoning seems to be a predictor of later metalinguistic awareness development. More broadly, this project sheds light on explaining the mechanisms behind FB development...
and identifies different pathways to FB understanding based on language status.

ACKNOWLEDGEMENTS

This work was completed as part of the first author’s doctoral dissertation submitted to the University of Florida. The research was supported by the Jacqueline Goldman Graduate Student Research Award. We want to thank Scott A. Miller, Lise Abrams, and H. Wind Cowles for their input on the project, as well as the anonymous reviewers for their thoughtful commentary on the manuscript. This work would not have been possible without the undergraduate research assistants at the Cognitive Development Lab at the University of Florida, the preschool administrators who opened their doors to our research team, and the parents and children who volunteered.

REFERENCES


