The Perceptions of Bilingualism scales: Development and validation using item response theory

Sarah Surrain[1*] and Gigi Luk[2]

[1] Harvard Graduate School of Education, Cambridge, Massachusetts, United States of America

[2] Department of Educational and Counselling Psychology, McGill University, Montreal, Quebec, Canada

* Corresponding author

E-mail: sarahsurrain@g.harvard.edu (SS)

Abstract

An increasing number of children in the U.S. and around the world are exposed to multiple languages, yet there is considerable variation in individual bilingual outcomes. Previous research has shown that factors such as input, usage, and language history can help explain this variation, but less is known about the role of attitudes towards bilingualism, and no instrument currently exists for measuring such attitudes. Guided by previous theories on language perceptions, we describe two new scales developed to measure general perceptions of the value of bilingualism (study 1: Perceptions of Bilingualism, PoB) and parental perceptions of the value of bilingualism for one's child (study 2: Perceptions of Bilingualism for child, PoB+) in the United States. We use factor analysis and Item Response Theory (IRT) to test the reliability, dimensionality, and individual item contributions of each scale using a national online sample of 422 adults (study 1) and a subsample of 321 parents (study 2). The final 10 and 8-item scales demonstrate internal reliability, unidimensionality, and precision of the intended construct to be measured. We report associations between scale scores and demographic characteristics and discuss how an IRT approach can complement classical approaches to attitude scale development. The PoB and PoB+ are useful tools to explore the association between social perceptions of bilingualism and language use for oneself and for one's child.

Introduction

Around the world, migration and globalization are enabling more people to learn more than one language, especially children who are first- or second-generation immigrants. In the United States, an estimated 32% of children under 18 are exposed to a language other than English at home (1), and this estimate does not include children who speak English at home but are exposed to additional languages in school. However, exposure alone does not guarantee

acquisition of languages other than English in the U.S. (see 2 for a recent review on immigrant children). In fact, young children whose families speak a minority language at home are at risk of losing their first language after beginning school in the majority language (3–5). Previous research has found that differences in children's bilingual outcomes are associated with proximal contextual factors such as the quantity and quality of input in each language, active use of each language, and the history of language exposure and acquisition (6–8). Less is known about the role of more distal contextual factors such as attitudes towards bilingualism in the society and parents' perceptions of the value of bilingualism for their child. One challenge to understanding the effect of these factors is that while many research tools have been developed to measure bilingual language experience and acquisition history, few tools have been developed to measure attitudes towards bilingualism.

Theoretical models have long posited that in settings where children hear more than one language, parental attitudes towards early bilingualism are an important factor in the degree to which children acquire each language (9–12). However, these models have not been tested empirically as no validated instrument exists for measuring the degree to which parents see bilingualism as valuable in the larger society and for their child's development. The two reported studies are first steps towards developing and validating such an instrument, which we call the Perceptions of Bilingualism Scales. These scales are intended as research tools to measure perceptions of the social value of bilingualism among adults from diverse social and linguistic backgrounds. The aim of this paper is to report on the process we used to develop and validate two scales: The Perceptions of Bilingualism (PoB) scale and the Perceptions of Bilingualism for One's Child (PoB+) scale.

Literature Review

We first provide a brief overview of previous instruments designed to measure language attitudes. Second, we discuss the potential contribution of a new instrument that can be used to measure perceptions of the value of bilingualism in the U.S., particularly among parents of young children. Third, we preview the method used to refine and test the psychometric properties of the scale items – Item Response Theory. In the literature reviewed below, the terms *attitude*, *belief*, and *perception* are used. We follow the terms used in each source, but we intentionally chose the term *perception* for our scales. The reason for this was that while *attitude* can be defined as "a disposition to respond favorably or unfavorably to an object, person, institution, or event" (13), *perception* highlights the process of interpreting and making meaning based on one's own life experiences and context. In our discussion, we use the term *belief* to refer to specific cognitive judgments that may represent components of an individual's overall perceptions of bilingualism, as in Rosenberg and Hovland's (14) tripartite model of attitude.

Measuring language attitudes in the U.S. context

A rich body of literature has found language attitudes to be key predictors and outcomes of successful second language learning (15,16) as well as a factor in explaining first language attrition among adults (17,18). Most of these studies focused on measuring attitudes towards particular languages and language speakers, such as French and English in Canada (19,20), and Welsh and English in the United Kingdom (21–24). This work has used innovative approaches such as the Matched Guise technique to detect implicit biases towards particular language groups (20) and attitude scales designed to investigate whether language learners are motivated by instrumental or integrative goals (15,19). However, very little psychological research on language attitudes has been conducted in the U.S. context.

The U.S. is characterized by vast linguistic diversity, a globally high-status majority language (English), and relatively rapid linguistic assimilation for new immigrants. The U.S. is ranked fifth in terms of total number of languages spoken and first in terms of number of immigrant languages spoken (25). From 1980 to 2010, the number of U.S. residents who reported speaking a language other than English at home has grown from 23.1 to 59.5 million (26). U.S. educational policies tend to prioritize children's rapid acquisition of English, with much less attention paid to the maintenance of minority languages (27). The use of children's native languages for instruction in public schools has been hotly debated and even legally restricted in some states (28) and linguistic diversity has been seen by some as a threat to national unity (29,30). At the same time, a growing awareness of the cognitive, economic, and social benefits of bilingualism has fueled interest in dual language programs among English-speaking families (31). Against this backdrop, attitudes towards bilingualism in the U.S. are highly politicized and polarized (32).

Accordingly, U.S.-based research in this area has tended to focus on public policy and education. Studies in this vein have examined teachers' attitudes towards language diversity, parents' views of bilingual education, and language minority parents' views on transmitting their native language to their children. For example, Byrnes and colleagues (33,34) developed the Language Attitudes of Teachers Scale (LATS) and found that more positive attitudes towards linguistic diversity was observed among teachers who lived in linguistically diverse areas, held a graduate degree, and had experience with linguistically diverse students. A few quantitative studies have examined parents' attitudes towards bilingual education and dual language learning. These have mainly sampled parents who have a child already enrolled in a particular bilingual program (35,36) or parents from a particular geographic context or language background (37,38).

While these studies used questionnaires with items meant to tap into different dimensions of parents' attitudes and beliefs about bilingual education, none developed and validated a scale that could be used to measure perceptions of the value of bilingualism as a unitary construct.

Several qualitative studies have explored language minority parents' views on their child's maintenance of languages other than English (39–41), but little quantitative work exists. One exception is Mucherah's (42) scale measuring immigrant parents' perceptions of their native language. Mucherah found that most of the 208 immigrant parents she surveyed strongly desired that their child learn and use both English and their native language, but they were also concerned that native language use might have negative consequences on learning and peer relationships. This study was limited to immigrants to the U.S. who were proficient in written English and its focus was on perceptions of immigrants' mother tongue versus English, not their perceptions of the value of bilingualism more generally. One scale that did attempt to measure general perceptions of bilingualism was Baker's Attitude to Bilingualism scale (21). Using latent variable analysis with data from 797 Welsh adolescents, Baker found that most of the 25 Attitude to Bilingualism items loaded onto a single factor and were conceptually distinct from Attitudes to Welsh. These results suggest that attitudes towards bilingualism may be a distinct, unitary construct, but they have yet to be replicated in different populations and language contexts (see Pieras-Guasp, 2002 for one exception).

In sum, there remains a need for a brief instrument that can be used with parents from diverse language, immigration, and regional backgrounds to gauge their perception of the value of bilingualism in the society more broadly and for their child more specifically. Importantly, these may be two separate but related constructs, with how one sees the value of bilingualism in the society being distinct but predictive of what one wants for one's child. Given the context

specificity of bilingualism, we situated our investigation in the U.S. with the goal of providing insight into this particular sociolinguistic context.

The theoretical importance of measuring perceptions of bilingualism

An instrument that captures attitudes towards bilingualism would allow us to test theoretical models of bilingual language acquisition and minority language maintenance that feature attitudes as a key predictive factor. For example, De Houwer (9,10,44) theorizes that in language contact settings, parental attitudes drive their choices regarding which language(s) to use with their child and the strategies they employ to encourage or discourage their child's use of each language. Pearson's (12) input-proficiency-use cycle also assigns an important role to attitudes. She posits a self-reinforcing system in which when children use more of a particular language, this language choice elicits more input from adults and peers in that language, leading to greater proficiency, which in turn leads to greater use, and so on. In Pearson's model the attitudes held by parents, siblings and peers serve to promote or inhibit children's use of a language, with ripple effects on input and proficiency. A third model—Hamers and Blanc's Social-Psychological Model of Bilingual Development (11,45)—highlights the importance of adults' *valorization*, or the degree to which they see each language as a tool that supports cognition and social interaction. Adults' valorization of each language is internalized by children, influencing their motivation to learn and use each language. This model also posits that using language for successful communication ignites positive feedback loops that promote more positive attitudes and in turn fuel more language learning (11).

A reliable measure of perceptions of bilingualism would complement ongoing efforts to develop instruments measuring bilingual experience (46–49). Given that bilingualism reflects the interaction between individuals and their environment, documenting perceptions contributes to

this effort by bringing in a sociolinguistic perspective. A relatively short, efficient scale capturing perceptions of the social value of bilingualism that can be embedded in a longer language experience survey could allow us to explore whether perceptions of bilingualism vary by individuals' social and language background and by their sociolinguistic context.

Item Response Theory and the current study

Building on previous work on language attitudes in general, the current study applies an Item Response Theory (IRT) framework to assess the psychometric properties of the proposed PoB and PoB+. IRT complements and extends classical test theory approaches by allowing for the estimation of item parameters that are theoretically invariant across populations (50). In other words, IRT allows for a fine-grained evaluation of the amount of information each item contributes at different levels of the latent trait (here, the degree to which bilingualism is perceived as valuable). In this way, IRT can help determine which items to drop, modify, or retain, as well as provide estimates of measurement error that are conditional on trait level. In other words, it can help us shorten a scale without sacrificing precision and determine whether the scale is better at differentiating between individuals at the upper or lower ranges of scale scores. IRT is particularly useful when precise scores representing a single latent construct are of interest. For this reason, IRT is well-suited to the task of designing two unidimensional scales that can each be summarized in a single score. To our knowledge, IRT has not previously been used in the development of a language attitudes scale. Given that previous work has suggested that attitude towards bilingualism is a unitary construct based on latent variable analysis (21), IRT provides a systematic way to consider each item and its contribution to the scale.

In the current study we develop two separate scales. The first is intended to measure perceptions of the value of bilingualism in U.S. society (PoB) for use with adults, including

parents. The second is intended to measure parents' perceptions of the value of bilingualism for their child (PoB+). To develop and validate the two scales, we administered an online survey with two different national samples of adults. The data from both samples were combined in order to achieve sufficient power to reliably estimate IRT item parameters (51,52). In Study 1, the full combined sample was used to evaluate the psychometric properties of the PoB scale. In Study 2, the subset of participants that were parents of children under 18 was used to evaluate the PoB+ scale.

Study 1

The purpose of Study 1 was to develop and validate a scale measuring the degree to which adults perceive bilingualism as valuable in the United States. We sought to develop a scale that could be used with adults from diverse language and social backgrounds across the U.S. The study aims to answer 1) Does the PoB scale measure a single unitary construct in the U.S. sample? 2) What are the psychometric properties of each item, especially each item's contribution to the efficiency and coherence of the PoB scale? 3) How are PoB scale scores related to individual respondents' characteristics such as language background, education, age, and gender?

Study 1 Methods

Participants and procedure. Data were collected via Qualtrics Panels and Amazon's Mechanical Turk (MTurk). Qualtrics Panels provide a data collection service for a fee and Mturk is a crowdsourcing platform that amasses an online temporary workforce for different services, including survey participation. First, we contracted with Qualtrics Panels to recruit adults aged 18 and older that approximated the U.S. population on the basis of education, race/ethnicity, and region. Individuals who had previously registered with Qualtrics Panels were recruited via email

from March 7, 2016 – April 27, 2016. Parents of toddlers exposed to Spanish and English were oversampled to ensure sufficient representation for this more difficult-to-reach group for PoB+ (see study 2). To ensure that participants were able to read and understand the survey in their preferred language (English or Spanish) we included three comprehension checks consisting of multiple-choice items asking for the synonym of a common word. To ensure that participants read each scale item carefully we included two attention checks asking that a particular response be selected. In this first wave, 61 respondents were excluded due to failing either a comprehension check (n = 14) or an attention check (n = 47), resulting in a final sample of 210 participants. Of these, 109 were parents of a child under 18, and 73 of these had a child under three years old. Participants received a $1.00 credit to be used for online shopping for completing the survey.

A second sample was recruited via MTurk, an online platform frequently used in psychological research. We chose this platform because MTurk respondents have been found to better represent the U.S. population in terms of region, age, and race than other online samples (53). In this wave, participation was limited to adult parents living in the U.S. with at least one child under seven. The survey was open from January 31 – March 18, 2017. A total of 322 MTurk respondents were excluded for reporting that they did not have a child under 7 (n = 268), did not live in the U.S. (n = 53), or were under 18 years old (n = 1). Another 90 were excluded for failing a comprehension check (n = 34) or an attention check (n = 56). The final Mturk sample consisted of 212 participants who passed all eligibility screeners, comprehension checks, and attention checks. Participants received $.50 to compensate their time. Institutional review board approval was obtained for both rounds of data collection.

In addition to the PoB and PoB+ scales described below, the survey included questions about language background, educational attainment, and 10 items from MacPhee's widely used Knowledge of Infant Development Index (54). Slightly more than three quarters of the sample (77%) spoke English as a first language, and 17% spoke Spanish as first language (or learned Spanish and English simultaneously). Collectively, 34 different languages were represented as native languages and additional languages in the entire sample (see Appendix A in the online supplementary materials at https://surrain.com/s/pob1 for a complete list). Although Spanish versions of the scales and questionnaire were available, only 4% of the respondents took the survey in Spanish. The median time participants took to complete the survey was 13 minutes.

As shown in Table 1, there were some differences in participant demographics by sample. Sample 2, recruited via MTurk, was younger, had more education, and included fewer Black and Hispanic respondents than Sample 1. These differences can be attributed to the fact that Sample 1 was designed to meet certain quotas for each education and race category, and Sample 2 was only restricted to parents of young children. As minimal changes were made to the items and procedure across the two rounds of data collection, the two samples were combined for all analyses.

Table 1. Participant Demographics for Samples 1 and 2.

	Sample 1 (n = 210)	Sample 2 (n = 212)	Combined (n = 422)	Difference between Samples 1 and 2
Age	42 (17)	33 (7)	37 (14)	$t(420) = 7.29$, $p < .001$
Female	64%	60%	62%	$X^2(1) = 0.68$, $p = .409$
Born in U.S.	86%	90%	88%	$X^2(1) = 1.54$, $p = .215$
L1 English, doesn't speak an L2	46%	43%	44%	$X^2(1) = 0.33$, $p = .564$
Region of U.S.				
West	26%	21%	24%	$X^2(3) = 3.10$, $p = .541$
South	35%	38%	36%	
Midwest	22%	21%	22%	
Northeast	16%	19%	18%	
Education				
Some high school or less	10%	0%	5%	$X^2(4) = 45.69$, $p < .001$
High school graduate	27%	13%	20%	

Some college or ass. Degree	27%	25%	26%	
College Graduate	26%	45%	36%	
Graduate School Degree	10%	17%	24%	
Ethnicity/Race				
White alone (not Hispanic)	50%	67%	58%	$X^2(4) = 21.65$, $p < .001$
Black	11%	4%	8%	
Asian or Pacific Islander	5%	8%	6%	
Hispanic	31%	17%	24%	
Mixed race or other	3%	4%	3%	

Development of the PoB scale. An initial set of 13 items was developed based on our review of the literature, cognitive interviews, and input from members of the research team in our lab who have worked with linguistically diverse populations across the lifespan. The initial set of items (see Table 2) covered perceptions of whether speaking multiple languages in the U.S. should be acknowledged, accommodated, rewarded and supported; whether speaking multiple languages in the U.S. is needed and valued; and whether speaking multiple languages incurs personal benefits and costs. Several items were adapted from Baker's Attitude to Bilingualism Scale (21) and Byrnes and Kiger's Language Attitudes of Teachers Scale (LATS; 33,34).

Table 2. Initial set of 13 PoB items with the means and standard deviations (n = 422).

		M	SD
PoB1	The ability to speak more than one language is highly valued in the United States.	4.77	1.30
PoB2	The United States should have more than one official language.	3.40	1.71
PoB3R	To be considered American, one should speak English. (Reversed)	2.63	1.55
PoB4	Languages in addition to English should be taught in public elementary schools.	4.69	1.31
PoB5R	Parents in the United States who don't speak English should learn English to help their children. (Reversed)	1.97	1.06
PoB6	People who speak more than one language should earn more money in the United States.	3.89	1.48
PoB7	To be successful in the United States you need to speak more than one language.	3.45	1.56
PoB8	Teachers, doctors, lawyers and police officers in the United States should speak a language in addition to English so they can communicate with the people they serve.	4.37	1.40
PoB9	Parents whose native language is not English should teach their native language to their children.	4.73	1.16
PoB10	Learning a second language helps a person think more creatively.	4.69	1.30
PoB11R	Learning a second language will negatively affect a person's first language. (Reversed)	4.77	1.44

PoB12	I wish I spoke another language (in addition to the language or languages I speak at this time).	4.80	1.36
PoB13	Speaking more than one language helps a person understand people from different cultural backgrounds.	4.77	1.30

Note: Items 3, 5, and 11 were reverse coded. These items were later dropped from the scale.

We chose to use a 6-point Likert scale from 1 (strongly disagree) to 6 (strongly agree) with no midpoint option elicit greater variability and discourage *satisficing,* or providing a response without expending the cognitive effort required to fully interpret and respond to each item (55,56). Each item was translated into Spanish by one of the two Spanish-English bilingual members of the research team and was checked for clarity and precision by the other member. Cognitive interviews were conducted by the first author in English (n = 6) and Spanish (n = 2) prior to administering online survey, and additional cognitive testing of the items was embedded in a subsequent qualitative study (n = 14) to further ensure that items were understood as intended in each language (57).

Analytic plan. There were three stages of data analysis. First, classical test theory and factor analysis were used to assess the internal consistency of the PoB scale and evaluate whether the scale was unidimensional. We elected to fit both an exploratory factor model and a single factor confirmatory model to evaluate model fit and correlations between item residuals. Following Hu and Bentler's (58) recommendations, we considered the following as indicators of good model fit: root mean square error of approximation (RMSEA) of less than .08, Bentler Comparative Fit Index (CFI) and Tucker Lewis Index (TLI) values greater than .90, and a standardized root mean square residual (SRMR) of less than .08. These steps are important because IRT is best suited to measuring a unidimensional construct precisely and relies on the assumption that item-level errors are uncorrelated (59).

Second, an IRT model was fit to estimate the discrimination and location parameters for each item on each scale. Because item responses were polytomous and ordinal, a graded response model (GRM) was fit with the following specifications:

$$P\left(Y_{pi} \geq k \middle| a_i, b_{ik}, \theta_p\right) = \frac{1}{\left(1 + \exp\left(-a_i\left(\theta_p - b_{ik}\right)\right)\right)};\ \theta_p \sim N(0,1)$$

In this model, *Y* is the response given by participant *p* on an item *i* that has *k* boundaries between response options. All items had six response options, resulting in 5 different boundaries (k = 1…5). The latent construct, hypothesized to be the degree to which bilingualism is seen as valuable in the society, is represented by theta (θ), which is constrained to a have mean of 0 and a standard deviation of 1. The discrimination parameter (a) represents the degree to which an item distinguishes between respondents with similar latent attitudes. The location parameter (b) tells us the theta score needed to have a 50% chance of scoring at or above each boundary. Based on results from the first two stages, three items were flagged for removal from the scale. Then, the overall precision and coherence of the scale were tested after removing these items from the scale. Finally, we examined the relation between theta scores estimated using the shortened scale and participant demographics using Pearson's correlations and multiple linear regression. All analyses were performed using Stata 15.1 statistical software.

Study 1 Results

Internal reliability and factor analysis. The initial 13 PoB items showed relatively strong internal reliability overall ($\alpha = .82$). The majority of the items had moderate to high item-total correlations, ranging from .59 to .75. However, two items had very low item-total correlations: PoB5R ("Parents in the United States who don't speak English should learn English to help their children", $r = .16$) and PoB11R ("Learning a second language will negatively affect a person's first language", $r = .18$). An exploratory factor analysis including the 13 items

revealed that 80% of the total variance was explained by a single factor, suggesting that the scale was predominantly unidimensional. The items with the highest factor loadings were PoB10 ("Learning a second language helps a person think more creatively") and PoB4 ("Languages in addition to English should be taught in public elementary schools"). The two items with the lowest loadings were PoB5R and PoB11R. Two additional items, PoB3R ("To be considered American, one should speak English") and PoB1 ("The ability to speak more than one language is highly valued in the United States") also had factor loadings of less than .40. The three items with the lowest correlations were all reverse-coded. See Appendix B, Table 1 for each item's item-total correlation and factor loading.

We then fit a single-factor confirmatory model with all 13 items. The model fit indices for this model were outside of the acceptable ranges (χ^2 *(df=65) = 501.23, RMSEA=.126, CFI=.765, TLI = .718, SRMR=.095).* An inspection of the correlation matrix of residuals from this model revealed large positive correlations between the errors of PoB3R and PoB5R and between PoB3R and PoB2. The errors of items PoB11R and PoB2 were strongly negatively correlated (see Appendix B, Table 2). These large correlations between the errors of the reverse-coded items and other items on the scale provide additional evidence that responses to these items may systematically covary with factors other than the construct of interest, which would violate the assumption of local independence required for IRT.

Item Response Theory. After finding the initial 13-item PoB scale to display adequate internal consistency and unidimensionality, we proceeded to stage two of our analysis. A graded response IRT model was fit for the full sample of 422 adults for the PoB scale. Table 3 displays the estimated item parameters for the PoB scale. Discrimination parameters indexing the degree to which each item provides information about the underlying trait ranged from .02 to 2.71. The

items providing the most information to scale scores estimated from this model were about the cognitive and social benefits of bilingualism (PoB10, PoB13) and whether bilingualism should be supported in schools (PoB4). The three reverse-coded items flagged in the previous stage stand out as providing very little information about the latent attitude being measured, with discrimination scores of <.5 The item about whether parents should learn English to help their children (PoB5R) was particularly uninformative (a = .02). This finding is understandable because the item was originally intended as an indicator of negative attitudes towards language minority parents with limited English skills, but it could also be interpreted as a marker of positive attitudes towards language minority parents' bilingual learning. Following the three reverse-coded items, PoB1 ("The ability to speak more than one language is highly valued in the United States") contributes the least information, even though on its face it is central to the construct. This item may need to be reworded in future iterations of the scale.

Table 3. PoB Item discrimination and location parameter estimates for the 13-item scale, estimated from a Graded Response Model (n = 422).

Item	Discrimination Parameter Estimates	Location Parameter Estimates				
	a	b1	b2	b3	b4	b5
PoB1	0.89	-4.85	-2.95	-2.23	-0.94	0.67
PoB2	1.56	-1.31	-0.57	-0.02	0.67	1.52
PoB3R	0.47	-1.79	0.35	2.28	3.80	5.83
PoB4	2.19	-2.31	-1.84	-1.27	-0.43	0.54
PoB5R	0.02	-18.48	49.35	141.99	188.77	212.83
PoB6	1.63	-2.01	-1.28	-0.50	0.44	1.40
PoB7	1.23	-1.95	-0.88	-0.01	0.93	1.95
PoB8	1.83	-2.40	-1.50	-0.97	-0.11	0.90
PoB9	1.86	-2.95	-2.17	-1.59	-0.38	0.71
PoB10	2.71	-2.14	-1.80	-1.22	-0.31	0.47
PoB11R	0.46	-6.82	-4.86	-3.57	-2.02	0.76
PoB12	1.82	-2.52	-1.89	-1.36	-0.60	0.35
PoB13	2.04	-3.20	-2.39	-1.81	-0.60	0.56

Note: All items have 6 score points and thus 5 threshold location parameters.

The location parameters indicate the probability of responding at or above a particular response category, given one's latent perception of bilingualism. For example, individuals near the mean scale score ($\theta = 0$) would be expected to "somewhat agree" with PoB7 ("To be successful in the United States you need to speak more than one language") but "agree" with PoB12 ("I wish I spoke another language"). Items with higher location parameters overall do a better job of discriminating between individuals with more positive attitudes, while items with lower location parameters better differentiate between individuals with more negative attitudes. We can see in Table 3 that the majority of the PoB items' parameter estimates fall between about -3 and 1, suggesting that the scale is performing best for individuals with scores in the middle to lower range. The location parameter estimates also reveal items that are not functioning optimally. For example, in addition to its very low discrimination parameter, item PoB5R has location parameters that fall well outside the range of our data.

The Category Characteristic Curves in Fig 1 provide a visual depiction of these parameter estimates by mapping the probability of a response in each category at each level of latent attitude. Here we can see that while some items show less overlap between the curves (e.g. PoB13, "Speaking more than one language helps a person understand people from different cultural backgrounds"), items such as PoB3R, PoB5R and PoB11R show considerable overlap between the categories, suggesting that these items are less useful.

Fig 1. Category Characteristic Curves for PoB items, estimated from a Graded Response Model (*n* = 422).

Another way to visually depict the amount of information contributed by each item across different levels of latent attitudes (theta score, θ) is by graphing the item information functions. Fig 2 illustrates how PoB10 provides the most information, particularly between theta scores of about -2 to 1. The Item PoB7 provides less information overall but discriminates better

than PoB10 at higher levels of theta. Items PoB5R, PoB11R, and PoB3R again stand out as contributing very little information across all levels of theta. Because IRT carries the assumption of local independence, eliminating low-performing items from our scale should increase the scale's efficiency but not substantially change the precision of our estimates from the smaller set of items. Fig 3 illustrates the test information functions (the sum of all item information functions across levels of theta) for the original 13-item scale and a 10-item scale. We can see that after dropping the three reverse-coded items, the total information, and thus the precision of our estimates, is virtually unchanged. Moreover, after removing these three items, the internal reliability of the scale is improved ($\alpha = .86$), and 99% of the variation is now explained by a single factor (see Appendix B, Table 1).

Fig 2. Item information functions estimated from a Graded Response Model (*n* = 422).

Fig 3. Test information functions estimated from a Graded Response Model (*n* = 422).

Correlations and regression analysis. In the third stage of our analysis, we explored the correlations between PoB scale scores and demographic information, including language background, years of education, age, and gender. Language background was treated dichotomously as having English as a first language and not speaking any other language (*n* = 197) in contrast to speaking a language other than English as a first or additional language (*n* = 235). Education was treated as a continuous variable by substituting each level of attainment with the approximate years of education completed (e.g. high school graduate = 12 years). We expected that participants who spoke a language other than English would view bilingualism more positively, but we did not have a prediction for how PoB scores would relate to education, age, and gender. Table 4 shows that as expected, PoB scale scores were significantly associated with language background, such that participants who spoke only English tended to score lower

than those who spoke a language other than English ($r(420) = -.39$, $p < .001$). Having more years of education ($r(420) = .20$, $p < .001$) and being female ($r(420) = .19$, $p < .001$) were significantly positively associated with PoB scale scores. Age was significantly negatively associated with PoB scale scores ($r(420) = -.30$, $p < .001$), but this may be at least partially due to the moderately large correlation between language background and age, such that older respondents were also more likely to also speak English as a first and only language.

Table 4. Correlations between PoB scale scores and demographic characteristics ($n = 422$).

	PoB	L1 Eng no L2	Years of education	Age
L1 English, no L2	-.39***			
Years of education	.20***	-.18***		
Age	-.30***	.28***	-.10*	
Female	.19***	-.09	-.04	-.09

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

By adding all of these predictors into a multiple linear regression model, we found that language background is consistently predictive of PoB scale score when controlling for education, age, and gender (see Table 5). In other words, respondents who had English as first language and had never learned another language scored almost .6 standard deviations lower, on average, than respondents who spoke a language other than English. Each predictor made a significant, independent contribution to PoB scale scores. We tested all interactions, and only the interaction between language background and age was significant when controlling for the other two variables. The presence of this significant interaction suggests that older respondents scored lower on the PoB than younger respondents, but this difference by age was less pronounced for respondents who spoke a language other than English (see Fig 4).

Fig 4. Predicted PoB scale scores as a function of age for prototypical female respondents with 15 years of education who speak English as a first and only language (solid line) compared to those who speak a language other than English (grey dotted line).

Table 5. PoB scale score as a function of language background, education, age and gender (n = 422).

	Model 1	Model 2	Model 3	Model 4	Model 5
L1 English, no L2	-0.742*** (0.085)	-0.697*** (0.086)	-0.593*** (0.087)	-0.568*** (0.086)	-0.571*** (.086)
Years of education		0.055** (0.019)	0.051** (0.019)	0.055** (0.019)	0.051** (0.019)
Age (centered)			-0.014*** (0.003)	-0.013*** (0.003)	-0.005 (0.005)
Female				0.299*** (0.084)	0.300*** (0.084)
L1 Eng x Age					-0.014* (0.006)
Constant	0.329*** (0.057)	-0.499 (0.293)	-.483 (0.287)	-.743 (0.292)	-0.648 (0.294)
R^2	.153	.169	.207	.230	.240
df_m	1	2	3	4	5
df_r	420	419	418	417	416
F	75.59	42.58	36.32	31.15	26.23

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Standard errors in parentheses. Age is centered at the mean of 37.

Study 1 Discussion

In Study 1 we found that the PoB scale was primarily unidimensional as a 13-item scale, but the three reverse-coded items stood out as potentially problematic. Classical approaches showed these items to be minimally correlated with the other items. The weak correlations and low factor loadings for these items could be due to one or more of the following: First, these items may pertain to a distinct construct or constructs. For example, the belief that bilingualism has linguistic costs (PoB11R) may be unrelated to perceiving bilingualism as socially valuable. Second, these items may have been be too complex or unclear to be interpreted as intended across respondents. Finally, reverse-coded items on scales often load onto a separate factor from non-reverse-coded items (60,61), which could be an artifact of measurement error or tendencies to acquiescence to agree/disagree statements (62).

Our IRT analysis provided more detailed information about the specific shortcomings of these items and how excluding them would affect the scale. We found that dropping these three

items did not decrease the precision of the scale and improved its internal reliability and coherence. IRT also revealed that our scale is better at differentiating between individuals in the middle to lower ranges of latent attitudes towards bilingualism and is less equipped to differentiate between individuals in the upper ranges of attitudes. Moving forward, it will be important to consider whether the 10 remaining PoB items adequately cover the construct we intend to measure. It may be that the 10-item scale captures variation in perceptions of the value of bilingualism, but falls short of capturing beliefs about the linguistic costs of bilingualism or the importance of English in the U.S.

Finally, as expected, individuals who spoke languages other than English scored higher on the PoB on average. Participants who had more education, were younger, or were female tended to hold more positive perceptions of bilingualism. However, differences by age were more pronounced for English monolinguals. Among those who spoke a language other than English, perceptions of the value of bilingualism were similarly high across different ages.

Study 2

The purpose of Study 2 was to develop and validate a scale measuring the degree to which parents perceive bilingualism as valuable for their child. Unlike earlier scales that have been used with specific samples of language minority parents (42) or parents of children enrolled in bilingual programs (35), the PoB+ scale was designed to be used with parents and other caregivers from both minority and majority language backgrounds and whose children were not necessarily learning an additional language at school.

As in Study 1, after examining the reliability and item-level properties of the PoB+ scale, we used pairwise correlation and multiple regression to begin exploring the relation between scale scores and other demographic variables. Of particular interest were the relations between

PoB+ scores, parents' language background, and parents' awareness of developmental milestones as measured by the Knowledge of Infant Development Index (KIDI; 54). Items from the KIDI have been found to predict parental language input and child language skills in monolingual settings (63,64). However, we wanted to know if parents' knowledge of child development was predictive of perceptions of bilingualism or whether these two scores were unrelated. Similar to the research questions in Study 1, we aimed to answer these questions: 1) Does the PoB+ scale measure a single latent construct or multiple subconstructs? 2) What are the psychometric properties of each item, and how does each item contribute to the efficiency and coherence of the PoB+ scale? and 3) How are PoB+ scale scores related to PoB scores, parents' language background, knowledge of child development, and other demographic characteristics?

Study 2 Methods

Participants and procedure. The participants for Study 2 comprised the subset of Study 1 participants with a child under 18. This included 52% of those recruited via Qualtrics Panels (n = 109), and 100% of those recruited via MTurk (n = 212), for a total of 321. As shown in Table 6, the demographic characteristics of each sample were similar. Across both samples, almost all respondents reported being the child's parent or stepparent, with a small number (4%) reporting that they were the child's grandparent or other relative. For simplicity, we will use "mother" to refer to all female caregivers and "father" to refer to male caregivers. Knowledge of Infant Development Index (KIDI) scores represent the sum of correct answers, with a maximum possible score of 10. The mean KIDI score was 4.6 for sample 1 and 5.5 for sample 2. The two samples were combined for all analyses.

Table 6. Summary of Participant Demographics for Study 2 (n = 321).

	Sample 1 (n = 109)	Sample 2 (n = 212)	Combined (n = 321)	Difference between Samples 1 and 2
Parent age	35 (11)	33 (7)	34 (9)	$t(319) = 2.60$, $p = .009$
Parent is female	67%	60%	62%	$X^2(1) = 1.53$, $p = .216$
Parent was born in U.S.	80%	90%	87%	$X^2(1) = 6.55$, $p = .010$
Parent years of education	14.5 (2.2)	15.3 (1.8)	15.1 (2.0)	$t(319) = -3.52$, $p = .009$
Parent Knowledge of Infant Development (KIDI) score	4.6 (2.0)	5.5 (2.1)	5.2 (2.1)	$t(319) = -3.86$, $p < .001$
Parent L1 English, doesn't speak L2	30%	43%	39%	$X^2(1) = 4.86$, $p = .028$
Language parent speaks to child				
All English	48%	52%	50%	$X^2(3) = 10.74$, $p = .013$
Mostly English	13%	22%	19%	
Half English, half another language	32%	24%	27%	
Mostly or only another language	7%	2%	4%	

Development of the PoB+ scale. An initial set of seven items were developed for the PoB+ scale. Four items asked whether bilingualism confers particular benefits to one's child such as school success (PoB+2), economic opportunities (PoB+4), cognitive control (PoB+6), and cultural competence (PoB+8). Two items ask whether oral skills (PoB+1) or written skills (PoB+3) in multiple languages are important for one's child. One reverse-coded item was meant to get at the potential costs of bilingualism for one's child (PoB+5R). An eighth item (PoB+7R, also reverse-coded) was added after the first round of data collection and was intended to get at whether bilingualism is required for success, or if learning only the socially dominant language is sufficient. This last item, along with items PoB+1, PoB+3, PoB+4 and PoB+5R are similar to items on Baker's 1992 scale that was developed in Wales but adapted to the U.S. context. As with the PoB scale, responses were on a 6-point Likert scale from strongly disagree to strongly agree. Table 7 includes all eight items along with the means and standard deviations for the combined sample. (See Appendix B, Fig 1 for item score histograms).

Table 7. Perceptions of Bilingualism for one's child (PoB+) scale.

		M	SD
PoB+1	It is important for my child to SPEAK more than one language.	4.65	1.27
PoB+2	Speaking more than one language will help my child succeed in school in the long term.	4.83	1.18
PoB+3	It is important for my child to learn to READ and WRITE more than one language.	4.60	1.31
PoB+4	Speaking more than one language will help my child compete in the job market.	5.04	1.08
PoB+5R	My child will be confused if he or she learns two languages at the same time. (reversed)	4.45	1.53
PoB+6	Speaking more than one language will help my child become a stronger thinker.	4.88	1.18
PoB+7R	To be successful, the ONLY language my child needs to speak well is English. (reversed)	3.82	1.49
PoB+8	Speaking more than one language will help my child understand people from different cultural backgrounds.	5.05	1.05

Analytic plan. First, we checked the internal consistency and factor structure of the PoB+ scale using classical approaches. As in Study 1, we fit both an EFA and a CFA to evaluate dimensionality and IRT assumptions of local independence of item errors. Second, we applied a graded response IRT model to examine the parameters of each item and we explored whether excluding the lowest-performing items would improve the scale. Third, we looked at correlations between theta scores and other variables of interest. Finally, we fit a series of regression models to look at how PoB+ scores varied as a function of the respondents' language background, education, and gender.

Study 2 Results

Internal reliability and factor analysis. The 8-item PoB+ scale showed strong internal reliability ($\alpha = .88$). The item-total correlations ranged from .52 to .85 with a median of .85 (see Appendix C, Table 1). As in Study 1, the items with the weakest correlations with the total were the two reverse-coded items: PoB+5R ("My child will be confused if he or she learns two languages at the same time", $r(319) = .52$) and PoB+7R ("To be successful, the ONLY language my child needs to speak well is English", $r(210) = .62$). (Note the lower degrees of freedom for

this last item, as it was added to the PoB+ scale after the first wave of data collection.) An exploratory factor analysis using all 8 items showed the scale to be unidimensional; one factor explained 99% of the total variance. The items that were most representative of this factor were PoB+1 ("It is important for my child to SPEAK more than one language"), PoB+6 ("Speaking more than one language will help my child become a stronger thinker") and PoB+2 ("Speaking more than one language will help my child succeed in school in the long term") with factor loadings of .84, .82. and .81, respectively. Though the two reverse-coded items had the lowest loadings (.42 and .47), none of the loadings were below .40.

We then fit a one-factor confirmatory model and found that the overall fit was borderline acceptable (χ^2 *(df=20) = 89.50, RMSEA=.128, CFI=.922, TLI = .891, SRMR=.055).* The correlation matrix of residuals revealed a moderate positive correlation between the errors of items PoB+1 ("It is important for my child to SPEAK more than one language") and PoB+3 ("It is important for my child to READ and WRITE more than one language") ($r(210) = .21$) and between the errors of the two reverse-coded items ($r(210) = .45$) (see Appendix C, Table 2). Modification indices suggested that the greatest improvement to model fit could be achieved by allowing the errors for PoB+1 and PoB+3 to correlate. With this modification, indicators of global fit were significantly improved (χ^2 *(df=19) = 44.01, RMSEA=.079, CFI=.972, TLI = .959, SRMR=.047).* This suggests that in this sample, PoB+1 and PoB+3 may be contributing redundant information and thus exerting undue influence on our estimate of the latent attitude. In our IRT analysis below we explore how removing one of these items impacts parameter estimates and scale scores.

Item Response Theory. After finding that the PoB+ scale was clearly unidimensional, we proceeded to fit a graded response IRT model for the 321 parents for the PoB+ scale. Table 8

displays the discrimination and location parameter estimates for the PoB+ scale. Overall, the discrimination parameters are considerably higher than those for the PoB scale. The items that are most effective at distinguishing between participants with similar latent attitudes are PoB+1 ("It is important for my child to SPEAK more than one language") and PoB+3 ("It is important for my child to learn to READ and WRITE more than one language"). Consistent with our factor analysis results, the two items with the lowest discrimination parameter were PoB+5R (the reverse of "My child will be confused if he or she learns two languages at the same time") and PoB+7R (the reverse of "To be successful, the ONLY language my child needs to speak well is English"). These two reverse-coded items also stand out in Fig 5 as having flatter, overlapping Category Characteristic Curves compared to better-performing items such as PoB+1.

Table 8. PoB+ Item discrimination and location parameter estimates, estimated from a Graded Response Model (n = 321).

Item	Discrimination Parameter Estimates	Location Parameter Estimates				
	a	b1	b2	b3	b4	b5
PoB+1	3.72	-2.22	-1.63	-1.02	-0.26	0.45
PoB+2	3.31	-2.19	-1.76	-1.34	-0.58	0.42
PoB+3	3.27	-2.22	-1.55	-0.94	-0.28	0.52
PoB+4	2.92	-2.81	-2.01	-1.63	-0.73	0.20
PoB+5R	0.90	-3.66	-2.13	-1.47	-0.65	0.95
PoB+6	3.20	-2.37	-1.75	-1.38	-0.63	0.37
PoB+7R	1.24	-2.61	-1.54	-0.43	0.35	1.48
PoB+8	3.07	-2.66	-2.04	-1.64	-0.78	0.21

Note: All items have 6 score points and thus 5 threshold location parameters.

Fig 5. Category Characteristic Curves for all 8 PoB+ items, estimated from a Graded Response Model (n = 321).

As in Study 1, we compared the contribution of each item at each level of theta by inspecting the Item Information Curves (see Fig 6). Six of the eight PoB+ items are providing a great deal of information, particularly between -3 and 2 on the theta scale. The two reverse-

coded items are providing less information overall but do provide slightly more information than other items at theta values of about 2. Fig 7 shows the test information functions (the sums of the item information functions) for the 8, 7, and 6-item versions of the scale. Here, we can see that reducing the scale to six items by dropping items PoB+5R and PoB+7R would not substantially diminish the precision of our estimates. However, these two items were of substantive interest based on previous research and theory. Rather than dropping these two items completely, an additional subscale with these and other items that tap into parents' concerns about potential costs of bilingual development may be warranted in future iterations of this tool.

Fig 6. Item Information Curves for all 8 PoB+ items, estimated from a Graded Response Model (*n* = 321).

Fig 7. Test information functions for 8, 7 and 6 items, estimated from a Graded Response Model (*n* = 321).

We then addressed the potential violation of local independence raised by the positively correlated errors of items PoB+1 and PoB+3. These two items have similar wording and relatively high discrimination parameter estimates in the 8-item IRT model (a = 3.72 and 3.27, respectively). We asked whether these items may be providing redundant information, which could inflate their discrimination estimates. To assess whether our IRT model was robust to this possibility, we explored how excluding PoB+3 would affect our estimates and scale scores (59). We found that by fitting a graded response model with the remaining seven items, the range of discrimination parameters is largely unchanged (.95 - 3.52 instead of .90 - 3.72). The ranking of items by discrimination parameter shifts slightly (PoB+1 moves to fifth place and PoB+6 moves to first place). The mean difference in location parameters between the two models is .04, with three quarters of the location parameters shifting by more than .01. However, the overall scale

564 scores estimated by each model are highly correlated (r = .988), suggesting that dropping PoB+3
565 does not substantially change our estimates of latent attitude.

566 **Correlations and regression analysis.** We examined the correlations between PoB+
567 scale scores and other variables as a preliminary look at convergent and discriminant validity.
568 We expected that PoB+ scale scores would be significantly positively correlated with PoB scale
569 scores and negatively correlated with speaking only English, but we did not make a prediction
570 about their association with knowledge of infant development.

571 Table 9 shows that PoB+ scores were strongly correlated with PoB scores (r(319) = .82m
572 p<.001). Parents who spoke only English scored significantly lower on the PoB+ (r(319) = -.41,
573 p <.001). Mothers tended to score higher (r(319) = .19, p <.001). Parents with more years of
574 education scored slightly higher, on average (r(319) = .13, p = .016). Unlike in Study 1, the
575 correlation with age was non-significant, which may be due to the more restricted age range
576 among parents compared to among all adults. Finally, knowledge of infant development and
577 PoB+ scores were uncorrelated (r(319) = .01, p = .794*)*, suggesting that perceptions of the value
578 of bilingualism for one's child may be unrelated to parents' understanding of young children's
579 development.

Table 9. Correlations between PoB+ scale scores and demographic characteristics (n = 321).

	PoB+	PoB	L1 Eng no L2	Years of education	Age	Female
PoB scale score	.82***					
L1 English, no L2	-.41***	-.33***				
Years of education	.13*	.09	-.11*			
Age	-.09	-.10	.18**	.03		
Female	.19***	.22	-.10	-.10	-.02	
KIDI score	.01	.00	.14*	.11	.08	.18**

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

580

We then fit a series of multiple linear regression models with PoB+ scale score as the dependent variable. Our final model (Table 10, Model 5) included main effects of language background, education, and gender, as well as interactions between language background and gender and language background and education. Age and KIDI scores were not included in the model, as they were not found to be significant predictors of PoB+ scores. As with PoB scores, language background was consistently predictive of PoB+ scores, with parents who spoke only English scoring lower on average than parents who spoke languages in addition to English. However, the relation between language background and score depended on gender and years of education. Among parents who spoke a language other than English, mothers scored much higher than fathers, on average, but parents scored similarly across all levels of education. Among English monolingual parents, mothers and fathers scored similarly but varied by education, with more years of education predictive of higher PoB+ scores. Fig 8 illustrates the interaction between language background and education.

Table 10. PoB+ scale score as a function of language background, education, and gender (n = 321).

	Model 1	Model 2	Model 3	Model 4	Model 5
L1 English, no L2	-0.802*** (.099)	-0.782*** (.099)	-0.748*** (.099)	-0.423*** (.154)	-1.904*** (.745)
Years of education		0.042 (.024)	0.051* (.024)	0.047* (.024)	0.003** (.032)
Female			0.318** (.099)	0.537*** (.127)	0.526*** (.126)
L1 Eng x Female				-0.542** (.198)	-0.489* (.199)
L1 Eng x Education					0.097* (.048)
Constant	0.303*** (.062)	-0.344 (.374)	-0.687 (.384)	-0.768* (.381)	-0.092 (.505)
R^2	.170	.178	.204	.223	.233
df_m	1	2	3	4	5
df_r	319	318	317	316	315
F	65.47	34.48	27.11	22.61	19.09

Fig 8. Predicted PoB+ scale scores as a function of years of education for prototypical mothers who speak English as a first and only language (solid line) compared to those who speak a language other than English (grey dotted line).

Study 2 Discussion

In Study 2 we used a subset of the sample from Study 1 to develop and validate a scale to measure parents' perceptions of the value of bilingualism for their child. We found that all 8 items loaded onto a single factor, suggesting that the PoB+ scale measured a single unidimensional construct. Two reverse-coded items stood out as having relatively weaker correlations with the total score, though their factor loadings were not so low as to be problematic.

The results from our IRT graded response model suggested that most of the items performed well, and that the scale had greater precision among parents in the average to below-average range of latent attitude. The two reverse-coded items that asked about the costs of bilingualism contributed less information overall but contributed slightly more than other items in the upper ranges of latent attitude scores. This may be because some parents feel that the cost of bilingualism is an inevitable but transient developmental phase that precedes subsequent benefits and opportunities. Because the content of these items is of substantive interest, we opted to retain all 8 items. However, we acknowledge that in future iterations of this scale we may want to add more items pertaining to concerns about the costs of bilingualism for one's child, which the current PoB+ may not adequately account for.

As in Study 1, speaking a language other than English was positively associated with positive perceptions of the value of bilingualism for one's child, as were years of education and being female. However, parents' knowledge of children's early milestones was not related to PoB+ scores, suggesting that perceptions of the value of bilingualism for one's child is not simply a biproduct of having a greater understanding of how children develop. Interestingly, we found that among parents who spoke a language other than English, mothers scored significantly higher than fathers. Among parents who spoke only English, gender mattered much less but

education mattered more: mothers and fathers with more years of education had more positive perceptions of the value of bilingualism for their child than those with fewer years of education.

General Discussion

Parental language attitudes are thought to play a key role in parents' language use with their child and the language skills that their child develops. In settings where multiple languages are used, parents' views of the importance of bilingualism in the larger society and for their child may help explain variability in child outcomes in each language. In the U.S., where exposure to multiple languages is seen as desirable by some and problematic by others, attitudes towards bilingualism may be particularly variable and influential. In this paper we report on the development and validation of two scales designed to measure perceptions of the value of bilingualism. While previous efforts have tended to focus on particular language communities, this is the first to develop and test such an instrument with a diverse sample of adults – both parents and non-parents, monolingual and bilingual – from across the U.S. Additionally, we go beyond classical approaches and apply item response theory to evaluate the contribution of each item in each of the two scales.

In study 1 we found that the PoB scale – intended to measure general perceptions of the value of bilingualism in the U.S. – performed well after excluding three of the initial 13 items. We found that respondents with more positive attitudes on the final 10-item scale were more likely to speak a language other than English, be female, have more education, and be younger. The observed age effect could be due to generational differences but could also reflect how younger participants were more likely to be bilingual parents of toddlers in this particular sample. In Study 2, we found that the 8-item PoB+ scale – intended to measure parents'

perceptions of the value of bilingualism for their child – was even more internally consistent and unidimensional than the PoB scale. As in study 1, more positive perceptions of bilingualism were observed among respondents who spoke a language other than English, had more education, and were female. Interestingly, knowledge of infant development was unrelated to perceptions of the value of bilingualism for one's child, suggesting that the latent attitude we are measuring is not driven by knowledge of early milestones in child development. In both studies, the respondent's language background was a strong predictor of PoB and PoB+ scale scores and language background moderated other effects. Speakers of languages other than English generally saw bilingualism as more valued and valuable, whereas English-only speakers' perceptions were more strongly associated with demographic features such as age and education level.

While classical approaches are an essential first step in determining the internal consistency and coherence of a scale, IRT can extend our analyses to develop attitudes scales that are both precise and efficient. In the studies reported here, we examined the discrimination parameters of each item and visually inspected the amount of information that each item contributed at different levels of the latent attitude to determine which items to retain and which to exclude without sacrificing measurement precision. The final PoB and PoB+ include 10 and 8 items, respectively. The scales are short enough to embed in a longer questionnaire on parents' language history and current usage while still capturing variability in attitudes. IRT also allows us to answer the question: Among what type of respondents do these scales differentiate best? By estimating the conditional standard error of measurement for different levels of the latent attitude score, we learned that both scales discriminate better between people in the lower and middle range of perceptions of the value of bilingualism. Finally, the item-level information that IRT generates can point us toward next steps in our scale development. Though IRT cannot directly

evaluate content validity—whether the scales actually measure the constructs that we intend them to measure—IRT can raise questions to be answered with future iterations of the scale. Below we discuss two questions of content validity that surfaced in our analyses.

The first question is whether the PoB and PoB+ scales adequately cover perceptions of the *consequences* of bilingualism, and if this is a component of one's perception of the value of bilingualism or if this is another dimension that requires a separate subscale. Items such as "Learning a second language will negatively affect a person's first language" on the PoB and "My child will be confused if he or she learns two languages at the same time" on the PoB+ sought to capture this, but these items were only weakly correlated with the rest of the items. Our IRT results indicated that the item regarding negative effects of learning a second language should not be included in the final PoB scale, and that the item regarding confusion due to learning two languages simultaneously provided the least information towards the latent attitude score on the final eight-item PoB+. These results are consistent with what Baker (21) found using factor analysis of his original Attitude to Bilingualism scale. Baker's item "children get confused when learning English and Welsh" had a factor loading of less than .3 on the same latent variable as the majority of his Attitude to Bilingualism items. The weak correlation between these items and the latent attitude towards bilingualism in both Baker's scale and our scales suggests that individuals can highly value bilingualism while also believing that children learning multiple languages may follow different developmental trajectories when compared to those learning only one. Future adaptations of the scales should include more items that capture this dimension to further explore this subconstruct.

The second question is whether parents perceive the importance of bilingualism as distinct from the importance of biliteracy. We intended for items 1 and 3 of the PoB+ scale to

capture two different ideas: the importance of one's child being able to *speak* multiple languages (oral language skills) and the importance of one's child being able to *read* and *write* in multiple languages (written language skills). While oral skills can be a foundation for literacy, literacy is not automatically acquired through naturalistic exposure (65,66). Past research suggests that language minority parents who value oral abilities in the minority language do not always view literacy in the minority language as equally important (41,67). However, in the current study we found that responses to these two items were highly correlated ($r(319) = .78$, $p < .001$) and had significantly positively correlated error terms in our CFA model. It is possible that parents perceived the two questions as asking essentially the same question, and thus they provided similar responses for both. This could be because they assumed that oral and written language skills are acquired and developed in a similar way, or because they were imagining their child learning an additional language in a school setting where written skills are emphasized alongside oral skills. If these two items provide redundant information to our IRT model, this could inflate our estimates of the amount of information provided by both items. Our IRT analysis showed that after excluding one of these items, PoB+ scale scores were largely unchanged. However, researchers using this scale may want to consider the population of parents under study and adapt or combine these items accordingly.

In sum, the two studies reported here describe the development and validation of two attitude scales intended to measure perceptions of the value of bilingualism in the society and for one's child. Using both classical approaches and Item Response Theory, we evaluate the internal consistency, precision, and efficiency of each scale. By drawing on a socio-economically and linguistically diverse sample of respondents from across the United States, we explore how perceptions of the value of bilingualism covary with demographic characteristics. The next steps

for this work include testing whether the perceptions measured by these scales are predictive of parents' language behaviors and their child's language outcomes, and adapting the scales for use in different sociolinguistic contexts outside the United States.

In the United States and around the world, the forces of migration and globalization increasingly bring languages into contact and create the potential (and need) for children to develop proficiency in multiple languages for success in school and life. Yet the same forces that produce multilingual contexts often threaten the intergenerational transmission of minority languages. While we know more about the impact of exposure to input in each language, parental perceptions of the value of bilingualism may also play a key role in whether children develop and maintain skills in more than one language. The ability to measure parents' perceptions would allow us to explore this phenomenon within and across different sociolinguistic contexts and populations provide empirical support for theories of bilingual language development.

Acknowledgements

We gratefully acknowledge Gladys Aguilar, Alexandra Chen, Dasha Maghooli, and So Yeon Shin for their contributions to the development of these scales. We would also like to Andrew Ho for his methodological guidance and Ruth Kircher for her helpful comments. Parts of the research reported in this manuscript were presented at the International Symposium on Bilingualism (June 2019). This project was funded by The Dean's Venture Fund at Harvard Graduate School of Education to the second author.

References

1. Child Trends. Dual Language Learners [Internet]. 2014 Nov [cited 2016 Feb 27] p. 1–22. Available from: http://www.childtrends.org/?indicators=dual-language-learners

2. Hoff E. Bilingual development in children of immigrant families. Child Dev Perspect. 2018 Jun;12(2):80–6.

3. Montrul S. Age of onset of bilingualism effects and availability of input in first language attrition. In: Nicoladis E, Montanari S (Linguist), editors. Bilingualism across the lifespan: factors moderating language proficiency. Washington, DC; Berlin: American Psychological Association: Walter de Gruyter GmbH; 2016. (Language and the Human Lifespan).

4. Portes A, Hao L. E pluribus unum: Bilingualism and loss of language in the second generation. Sociol Educ. 1998;71(4):269–94.

5. Wong Fillmore L. When learning a second language means losing the first. Early Child Res Q. 1991;6(3):323–46.

6. Bedore LM, Peña ED, Griffin ZM, Hixon JG. Effects of age of English exposure, current input/output, and grade on bilingual language performance. J Child Lang. 2016 May;43(3):687–706.

7. Hoff E, Core C, Place S, Rumiche R, Señor M, Parra M. Dual language exposure and early bilingual development. J Child Lang. 2012 Jan;39(01):1–27.

8. Song L, Tamis-LeMonda CS, Yoshikawa H, Kahana-Kalman R, Wu I. Language experiences and vocabulary development in Dominican and Mexican infants across the first 2 years. Dev Psychol. 2012 Jul;48(4):1106–23.

9. De Houwer A. Environmental factors in early bilingual development: The role of parental beliefs and attitudes. In: Extra G, Verhoeven L, editors. Bilingualism and migration. Berlin: Mouton de Gruyter; 1999. p. 75–96. (Studies on language acquisition ; 14).

10. De Houwer A. Harmonious bilingual development: Young families' well-being in language contact situations. Int J Biling. 2015 Apr 1;19(2):169–84.

11. Hamers JF, Blanc M. Bilinguality and bilingualism. 2nd ed. Cambridge; New York: Cambridge University Press; 2000. xiv+468.

12. Pearson BZ. Social factors in childhood bilingualism in the United States. Appl Psycholinguist [Internet]. 2007 Jul [cited 2017 Jan 24];28(3). Available from: /core/journals/applied-psycholinguistics/article/div-classtitlesocial-factors-in-childhood-bilingualism-in-the-united-statesdiv/CFC8ACE008981588CB28DF1FBF13EA5D

13. Ajzen I. Attitudes, personality and behavior. Milton Keynes: Open University Press; 1988. xiv+175. (Mapping social psychology).

14. Rosenberg MJ, Hovland CI. Cognitive, affective and behavioural components of attitudes. In: Rosenberg MJ, Hovland CI, editors. Attitude organization and change: An analysis of consistency among attitude components. New Haven: Yale Univer. Press; 1960. (Attitude organization and change).

15. Gardner RC. Social psychology and second language learning: the role of attitudes and motivation. London ; Baltimore, Md., U.S.A.: EArnold; 1985. xiv+208. (The Social psychology of language ; 4).

16. Gardner RC, MacIntyre PD. On the measurement of affective variables in second language learning. Lang Learn. 1993 Jun 1;43(2):157–94.

17. Ben-Rafael M, Schmid MS. Language attrition and ideology: Two groups of immigrants in Israel. In: Köpke B, Schmid MS, Keijzer M, Dostert S, editors. Language attrition: Theoretical perspectives. Philadelphia, PA: John Benjamins Publishing Company; 2007. p. 205–26. (Studies in bilingualism ; v. 33).

18. Cherciov M. Investigating the impact of attitude on first language attrition and second language acquisition from a Dynamic Systems Theory perspective. Int J Biling. 2013 Dec 1;17(6):716–33.

19. Gardner RC, Lambert WE. Attitudes and motivation in second-language learning. Rowley, Mass.: Newbury House Publishers; 1972. vi+313.

20. Lambert W, Hodgson R, Gardner R, Fillenbaum S. Evaluational reactions to spoken languages. J Abnorm Soc Psychol. 1960;60:44–51.

21. Baker C. Attitudes and language. Philadelphia: Multilingual Matters; 1992. 173 p. (Multilingual matters (Series) ; 83).

22. Jones WR. Attitude towards Welsh as a second language, a preliminary investigation. Br J Educ Psychol. 1949;19(1):44–52.

23. Jones WR. Attitude towards Welsh as a second language: A further investigation. Br J Educ Psychol. 1950;20(2):117–132.

24. Sharp D, Thomas B, Price E, Francis G, Davies I. Attitudes to Welsh and English in the schools of Wales. Macmillan Education Ltd, c/o Little Essex Street, London WC2, England; 1973.

25. Simons GF, Fennig CD, editors. Ethnologue: Languages of the world [Internet]. 20th ed. Dallas, TX: SIL International; 2017. Available from: http://www.ethnologue.com

26. Rumbaut RG, Massey DS. Immigration and language diversity in the United States. Daedalus. 2013;142(3):141–54.

27. García O, Kleifgen JA. Educating emergent bilinguals: Policies, programs, and practices for English language learners. New York: Teachers College Press; 2010. xi+180. (Language and literacy series (New York, N.Y.)).

28. Menken K. Restrictive language education policies and emergent bilingual youth: A perfect storm with imperfect outcomes. Theory Pract. 2013 Jul 3;52(3):160–8.

29. Hopkins DJ, Tran VC, Williamson AF. See no Spanish: language, local context, and attitudes toward immigration. Polit Groups Identities. 2014 Jan 2;2(1):35–51.

30. Petrovic JE. Balkanization, bilingualism, and comparisons of language situations at home and abroad. Biling Res J. 1997 Spring;21(2/3):233–54.

31. Adamy J. Dual-language classes for kids grow in popularity; in increasingly global economy, more parents seek leg up for their children through early immersion programs. Wall Street Journal (Online). 2016 Apr 1;n/a.

32. Schildkraut DJ. Press one for English: Language policy, public opinion, and American identity. Princeton, NJ: Princeton University Press; 2005. x+244.

33. Byrnes DA, Kiger G. Language attitudes of teachers scale (LATS). Educ Psychol Meas. 1994 Mar 1;54(1):227–31.

34. Byrnes DA, Kiger G, Lee Manning M. Teachers' attitudes about language diversity. Teach Teach Educ. 1997 Aug;13(6):637–44.

35. Cazabon M, Lambert WE, Hall G. Two-way bilingual education: A progress report on the Amigos program. 1993 Jan 1 [cited 2018 Aug 8]; Available from: https://escholarship.org/uc/item/633345b4

36. Craig BA. Parental attitudes toward bilingualism in a local two-way immersion program. Biling Res J Phila. 1996 Summer;20(3/4):383–410.

37. Farruggio P. Latino immigrant parents' views of bilingual education as a vehicle for heritage preservation. J Lat Educ. 2010 Jan 5;9(1):3–21.

38. Mancilla-Martinez J, Lesaux NK. Spanish-speaking parents' beliefs about their young children's learning and language development. NHSA Dialog [Internet]. 2014 Mar 14 [cited 2017 Jan 19];17(1). Available from: https://journals.uncc.edu/dialog/article/view/207

39. Cycyk LM, Hammer CS. Beliefs, values, and practices of Mexican immigrant families towards language and learning in toddlerhood: Setting the foundation for early childhood education. Early Child Res Q [Internet]. 2018 Dec 2 [cited 2018 Dec 8]; Available from: http://www.sciencedirect.com/science/article/pii/S0885200617301710

40. Kang H-S. Korean-immigrant parents' support of their American-born children's development and maintenance of the home language. Early Child Educ J. 2013 Nov;41(6):431–8.

838 41. Velázquez I. Intergenerational Spanish transmission in El Paso, Texas: Parental perceptions
839 of cost/benefit. Span Context. 2009 Apr 1;6(1):69–84.

840 42. Mucherah W. Immigrants' perceptions of their native language: Challenges to actual use
841 and maintenance. J Lang Identity Educ. 2008 Jul;7(3–4):188–205.

842 43. Pieras-Guasp F. Direct vs. indirect attitude measurement and the planning of Catalan in
843 Mallorca. Lang Probl Lang Plan. 2002 May 1;26(1):51–68.

844 44. De Houwer A. Bilingual First Language Acquisition. Bristol: Multilingual Matters; 2009.

845 45. Hamers JF, Blanc M. Towards a social-psychological model of bilingual development. J
846 Lang Soc Psychol. 1982;1(1):28–49.

847 46. Byers-Heinlein K, Schott E, Gonzalez-Barrero AM, Brouillard M, Dubé D, Jardak A, et al.
848 MAPLE: A multilingual approach to parent language estimates. 2018;

849 47. DeAnda S, Bosch L, Poulin-Dubois D, Zesiger P, Friend M. The language exposure
850 assessment tool: Quantifying language exposure in infants and children. J Speech Lang
851 Hear Res. 2016 Dec 1;59(6):1346–56.

852 48. Li P, Zhang F, Tsai E, Puls B. Language history questionnaire (LHQ 2.0): A new dynamic
853 web-based research tool. Biling Lang Cogn. 2014 Jul;17(03):673–680.

854 49. Marian V, Blumenfeld HK, Kaushanskaya M. The language experience and proficiency
855 questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. J
856 Speech Lang Hear Res. 2007 Aug;50(4):940–67.

857 50. Yen WN, Fitzpatrick AR. Item response theory. In: Brennan RL, National Council on
858 Measurement in Education, American Council on Education, editors. Educational
859 measurement. 4th ed. Westport, CT: Praeger Publishers; 2006. p. 111–8. (American
860 Council on Education/Praeger series on higher education).

861 51. De Ayala RJ. The influence of multidimensionality on the graded response model. Appl
862 Psychol Meas. 1994 Jun 1;18(2):155–70.

863 52. Reise SP, Yu J. Parameter recovery in the graded response model using MULTILOG. J
864 Educ Meas. 1990;27(2):133–44.

865 53. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: a new source of
866 inexpensive, yet high-quality, data? Perspect Psychol Sci. 2011 Jan 1;6(1):3–5.

867 54. MacPhee D. Knowledge of infant development inventory: Manual. Fort Collins CO:
868 Colorado State University; 2002.

869 55. Krosnick JA. Response strategies for coping with the cognitive demands of attitude
870 measures in surveys. Appl Cogn Psychol. 1991 Jun 5;5(3):213–36.

871 56. Krosnick JA, Fabrigar LR. Designing rating scales for effective measurement in surveys.
872 In: Lyberg L, Biemer M, Collins E, de Leeuw E, Dippo N, Schwarz N, et al., editors.
873 Survey measurement and process quality. New York: John Wiley; 1997. p. 141–164.

874 57. Surrain S. 'Spanish at home, English at school': how perceptions of bilingualism shape
875 family language policies among Spanish-speaking parents of preschoolers. Int J Biling
876 Educ Biling. 2018 Nov 26;0(0):1–15.

877 58. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis:
878 Conventional criteria versus new alternatives. Struct Equ Model Multidiscip J. 1999
879 Jan;6(1):1–55.

880 59. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire
881 development, evaluation, and refinement. Qual Life Res. 2007 Aug 1;16(1):5.

882 60. Colwell N, Gordon RA, Fujimoto K, Kaestner R, Korenman S. New evidence on the
883 validity of the Arnett Caregiver Interaction Scale: Results from the Early Childhood
884 Longitudinal Study-Birth Cohort. Early Child Res Q. 2013 Apr 1;28(2):218–33.

885 61. Duckworth AL, Peterson C, Matthews MD, Kelly DR. Grit: Perseverance and passion for
886 long-term goals. J Pers Soc Psychol. 2007 Jun;92(6):1087–101.

887 62. Krosnick JA, Judd CM, Wittenbrink B. The measurement of attitudes. In: Albarracin D,
888 Johnson BT, Zanna MP, editors. The handbook of attitudes. Mahwah, N. J.: Lawrence
889 Erlbaum Associates; 2005. p. 21–76.

890 63. Rowe ML. Child-directed speech: Relation to socioeconomic status, knowledge of child
891 development and child vocabulary skill. J Child Lang [Internet]. 2008 Feb [cited 2015 Dec
892 28];35(01). Available from:
893 http://www.journals.cambridge.org/abstract_S0305000907008343

894 64. Rowe ML, Denmark N, Harden BJ, Stapleton LM. The role of parent education and
895 parenting knowledge in children's language and literacy skills among White, Black, and
896 Latino families. Infant Child Dev. 2016;25:198–220.

897 65. Gough PB, Hillinger ML. Learning to read: An unnatural act. Bull Orton Soc.
898 1980;30:179–96.

899 66. Liberman IY, Shankweiler D, Liberman AM. The alphabetic principle and learning to read.
900 1989 [cited 2019 Jun 16]; Available from: https://eric.ed.gov/?id=ED427291

901 67. Lao C. Parents' Attitudes Toward Chinese-English Bilingual Education and Chinese-
902 Language Use. Biling Res J. 2004;28(1):99–121.

903

904

Open Practices Statement

The data used for all analyses are available via the Open Science Framework at

https://osf.io/yw7t2/

Online Supplementary Materials (available at https://surrain.com/s/pob1)

Appendix A.
Table 1. Number and % of sample who speak each of the 34 languages represented as either a first or additional language.

Appendix B.
Table 1. Item-total correlations and factor loadings for the 13-item and 10-item versions of the PoB scale (n = 422).
Fig 1. Item-score histograms for all 13 PoB items (n = 422).
Table 2. Residual matrix of a single-factor CFA model for all 13 PoB items (n = 422).
Fig 2. Factor Analysis Scree Plot for the 13-item PoB.

Appendix C.
Table 1. Item-total correlations and factor loadings for the 8-item PoB+ scale as well as the 7- and 6-item versions of the PoB+ scale (n = 321).
Fig 1. Item-score histograms for all 8 PoB+ items (n = 321; n = 212 for item PoB+7R).
Table 2. Residual matrix of a single-factor CFA model for all 8 PoB+ items (n = 321).
Fig 2. Factor Analysis Scree Plot for the 8-item PoB+.

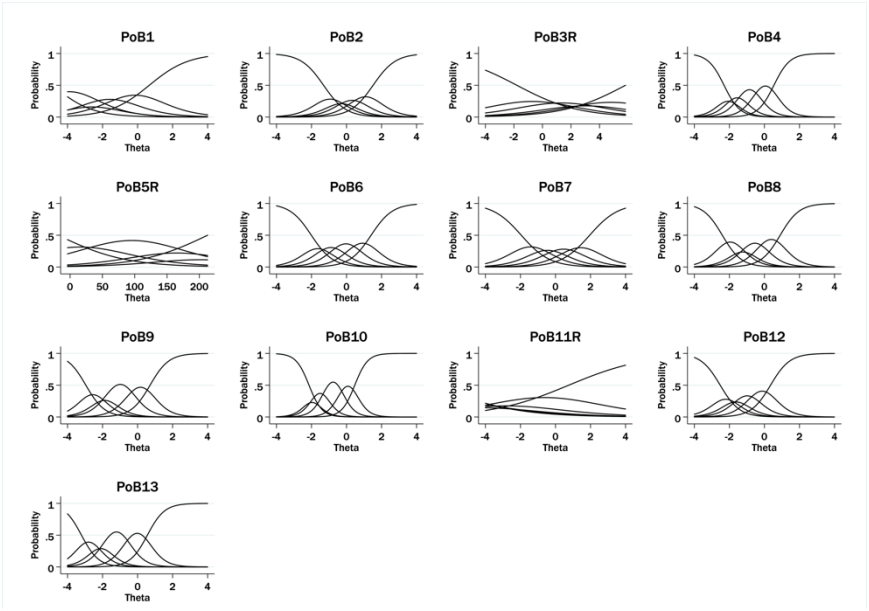

Figure 1. Category Characteristic Curves for PoB items, estimated from a Graded Response Model (n = 422).

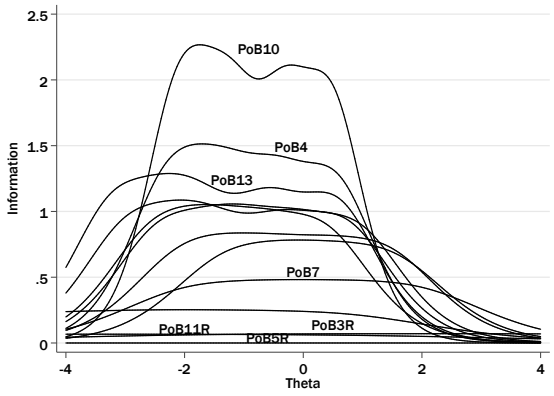

Figure 2. Item information functions estimated from a Graded Response Model (n = 422)

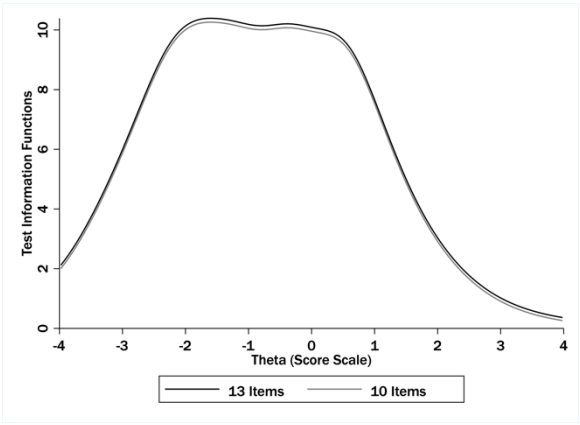

Figure 3. Test information functions estimated from a Graded Response Model (n = 422).

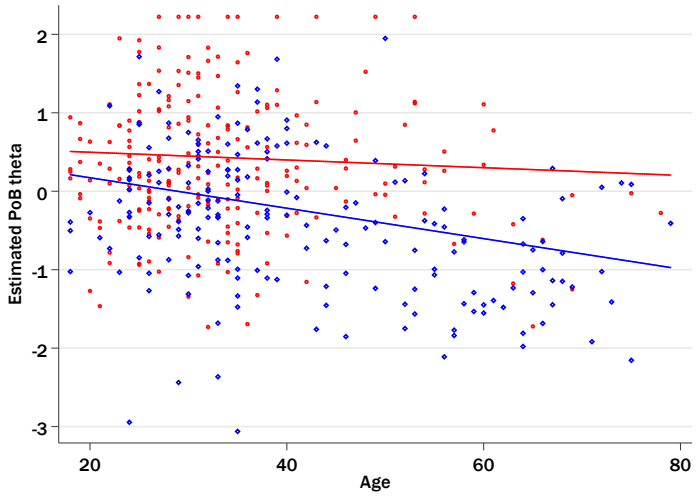

Figure 4. Predicted PoB scale scores as a function of age for prototypical female respondents with 15 years of education who speak English as a first and only language (solid line) compared to those who speak a language other than English (dotted grey line).

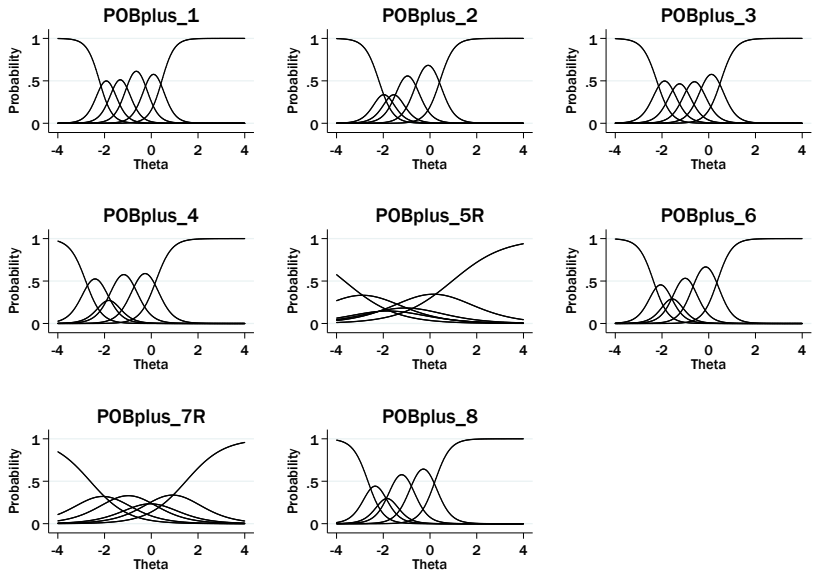

Figure 5. Category Characteristic Curves for all 8 PoB+ items, estimated from a Graded Response Model (n = 321).

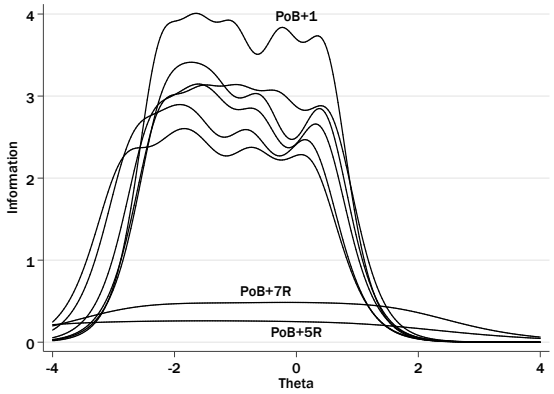

Figure 6. Item Information Curves for all 8 PoB+ items, estimated from a Graded Response Model (n = 321).

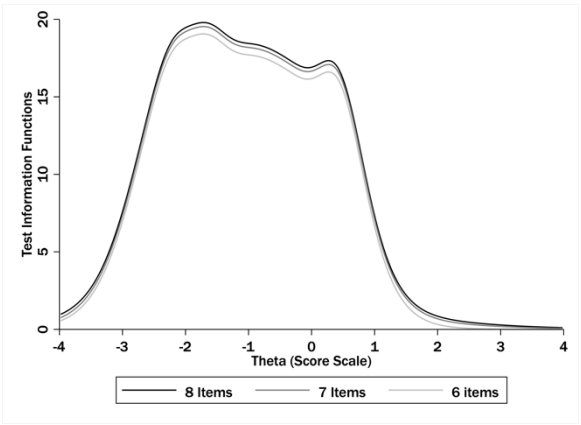

Figure 7. Test information functions for 8, 7 and 6 items, estimated from a Graded Response Model (n = 321).

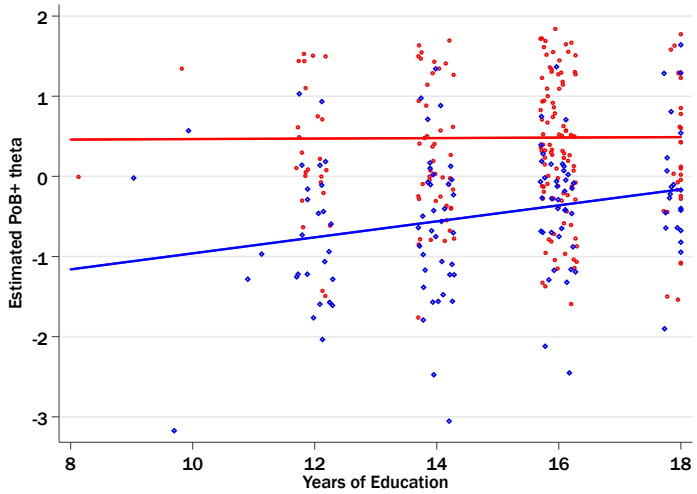

Figure 8. Predicted PoB+ scale scores as a function of years of education for prototypical mothers who speak a language other than English (red line) and English-speaking monolingual mothers (blue line).