Lost in Translation? Item Validity in Bilingual Political Surveys

Efrén O. Pérez Vanderbilt University

The dramatic increase in the U.S. Latino population in recent decades has spurred an equally dramatic rise in bilingual survey instruments used by scholars to gauge the political attitudes of this growing ethnic group. A key assumption behind these instruments is that English-language items tap the same political constructs as their Spanish-language analogs. This paper reports evidence which suggests that bilingual survey items may not always be comparable across linguistic groups. Using a variety of public opinion polls, I develop and test a series of multi-group measurement models showing that—net of measurement error—English- and Spanish-language survey items are not functionally equivalent. The paper discusses the implications of these findings for the development of future bilingual surveys, both in the United States and beyond, as well as the use of extant surveys for applied analyses of Latino political attitudes.

> "... all observers are not led by the same physical evidence to the same picture of the universe, unless their linguistic backgrounds are similar, or can in some way be calibrated."
>
> —Benjamin Whorf (1949)[1]

Are survey items equivalent across linguistic groups? Scholars of comparative politics have continuously wrestled with this challenge as they endeavor to measure political attitudes across linguistically varied nations (Ervin and Bower 1952–53; King et al. 2004; Stern 1948–1949). Today, this challenge has arisen in a context typically unencumbered by this concern: The United States of America. Sustained immigration flows into the United States have reconfigured the linguistic parameters of this nation. Most conspicuously, the United States has experienced an increase in Latinos, a growing ethnic group comprised of both English speakers and Spanish speakers (U.S. Census Bureau 2007). Scholars of racial and ethnic politics have relied on a deepening repository of survey data to measure Latinos' political attitudes (e.g., de la Garza, Garcia, and Falcon 1992; Nicholson, Pantoja, and Segura 2006; Sanchez 2006; Uhlaner and Garcia 2002). But since these survey instruments require translation from English into Spanish, it is reasonable to expect discrepancies in their performance. In particular, English-speaking Latinos may interpret survey questions differently than Spanish-speaking Latinos, thus raising concerns about the validity of bilingual survey items.

The default approach to ensuring cross-language validity in survey items is the "double blind" back translation method: A technique commonly used by researchers sampling English- and Spanish-speaking Latinos (Bobo et al. 1992–94: 232; Brislin 1970). According to this approach, a survey instrument is first developed by an individual or group of individuals with an English-speaking population in mind. This initial English-based instrument is then translated into Spanish, and then back into English, to ensure that both linguistic renditions arrive at a satisfactory level of equivalence.[2] Since the early

[1]The quote is taken from Jacobson, Kumata, and Gullahorn (1960).

[2]Not every bilingual survey is developed strictly according to these criteria. Some variants exist. For instance, de la Garza et al. (1989–90) used a focus group approach, whereby the instrument was developed separately in Spanish, and then compared and vetted with respect to the English version by a focus group comprised of six bilingual and college-educated individuals (two Mexicans, two Puerto Ricans, and two Cubans). This approach remains vulnerable to some of the same threats to validity as bilingual survey instruments developed along the more customary criteria. For instance, this approach risks administering survey items to a general population that does not share the characteristics (e.g., bilingual ability, high education) of the focus group participants.

The Journal of Politics, Vol. 71, No. 4, October 2009, Pp. 1530–1548 doi:10.1017/S0022381609990156
 ISSN 0022-3816

days of survey research, however, analysts have identified several pitfalls in this back translation method. One fundamental challenge is that by definition, language differences violate a strict sense of comparability across items because different words from different languages produce varying stimuli to survey respondents (Jacobson, Kumata, and Gullahorn 1960). This challenge is compounded by the availability of wide-ranging referents for the same concept in two or more languages, such that a direct translation of questions may fail to convey the intended meaning of the original items (Ervin and Bower 1952–53). Finally, although translated items may capture translators' comprehension of certain concepts, they may inadequately gauge these constructs as understood by a general population.

The design of cross-language survey items is thus fraught with pitfalls. But as with any research endeavor, the development of bilingual survey questions may introduce errors without systematically biasing these items (Andrews 1984; Asher 1974; DeShon 1998). Insofar as this is true, survey items may tap into the same political concept across different language groups in spite of the "noise" that inheres in any one measure. Emerging research suggests, however, that the language-of-interview systematically influences the answers provided by survey respondents, even after controlling for key demographic and political variables (Lee 2001). These persistent effects strongly suggest the need to further examine the psychometric properties of survey items administered to English- and Spanish-speaking Latinos. Such an inquiry can shed important light, not only on whether linguistic bias exists in these items, but also on the extent to which Latinos from different language backgrounds share similar political attitudes, beliefs, and values.

With these considerations in mind, this paper employs structural equation modeling (SEM) to ascertain the validity of bilingual survey items across English-speaking and Spanish-speaking Latinos (e.g., Byrne, Shavelson, and Muthén 1989; Chan et al. 2007; Diaz-Morales et al. 2006; Reise, Widaman, and PughReise, Widaman, and Pugh 1993; Saris and Andrews 1991; Spini 2003). This analytic approach has two main virtues in the context of bilingual survey data. First, it verifies the assumption that survey items capture the same phenomenon across English-speaking and Spanish-speaking Latinos, as is presumed when such items are used to compare these respondents across a given trait. And second, it gauges how well these items capture their intended concept by controlling for measurement error in specific items. Using a variety of data sets commonly used by scholars of Latino public opinion—such as the *Latino National Political Survey (1989–90)*, the *Multi-City Study of Urban Inequality (1992–94)*, and the *National Survey of Latinos (2004)*—I develop a series of multigroup measurement models to assess whether survey items gauge the same constructs in English as they do in Spanish. The investigation reveals that the cross-language validity of the analyzed items is weak. The paper discusses the implications of these findings for the study of language and public opinion in linguistically diverse societies, and it outlines specific steps to minimize this challenge in applied analysis as well as in the development of future bilingual survey items.[3]

What Do We Know About Item Validity in Bilingual Surveys?

Though bilingual survey instruments have gained currency with the increase in the U.S. Latino population, these types of instruments have a history stretching back decades to the early days of survey research in comparative contexts (e.g., Ervin and Bower 1952–53; Stern 1948–49). With the advent of public opinion surveys, scholars seized the opportunity to gauge citizens' attitudes in various countries, thus spurring the development of survey instruments into requisite languages (e.g., Almond and Verba 1963). These early efforts at cross-national survey research generated insight into some of the challenges that arise in the development of bilingual survey items. Today, these challenges jeopardize the validity of survey data on U.S. Latinos.

Developing valid survey items is a daunting task involving much trial-and-error. Researchers not only design measures that aim to capture a particular construct (i.e., attitude, belief, or value), they also empirically verify the performance of these items—an exercise which generally leads to a revision and retesting of one's measures (e.g., Chan et al. 2007; Davis and Brown 2002; Feshbach 1991; Kosterman and Feshbach 1989; Paunonen and Sampo 1998). Yet meeting this objective in a bilingual context introduces additional burdens to researchers. These burdens can

[3]Though this paper is on bilingual surveys administered to Latinos, the same concern with validity applies to other ethnic communities which vary by language, such as Asian-Americans.

be traced to one source: the need to achieve *functional equivalence*—that is, the same meaning—across two or more linguistic versions of survey items (Jacobson, Kumata, and Gullahorn 1960).

This objective is usually accomplished via the back translation method. According to this approach, a small set of individuals translates a survey instrument between two languages in order to ensure functional equivalence between both versions (Brislin 1970). Several issues typically arise in this method, however, thereby introducing some slippage between the different language versions of an instrument. One such challenge is the difficulty in ensuring that a translated item captures its intended meaning, or as Jacobson, Kumata, and Gullahorn (1960) note, that a balance between literal and functional equivalence is achieved. Although it is straightforward enough to directly translate words from one language into another, the tougher challenge resides in making sure that a translated item communicates its point clearly and precisely. For instance, Bower and Ervin (1952–53, 597) note that in a an instrument translated from English into Korean, a question probing *expectations* produced responses that could only be explained by the fact that *hope* was included in the Korean word for *expectations*. Similarly, Stern (1948–49, 713–14) recounts how in an early multilingual survey of Europe, respondents in one nation overreported their ownership of washing machines because they understood the question as including nonelectric, manually operated devices, even though the question was originally designed to gauge ownership of the electric variety. More recently, in an effort to use English-based analogs to develop Chinese-language items measuring depression, Chan et al. (2007) discovered that certain items held different meanings for Chinese subjects, which in turn deteriorated the validity of their measures.

The above challenges are compounded by the varying range of referents available for the same word or concept in a language, as well as the tension between words that contain different affective valences (Erwin and Bower 1952–53). For instance, what is the most appropriate way to translate the term African American into Spanish in a survey gauging Latino attitudes toward this racial group? The term *Afro-Americano* is the literal equivalent of African American, yet the suffix *–Americano* may produce confusion, as some respondents may infer that the term includes blacks in the Americas, rather than just blacks in the United States. The alternative is to use a term that refers to the skin color of African Americans, such as *negro*, *prieto*, or *moreno*. Yet the choice of any of these three is not innocuous, as the first two arguably contain a more negative valence than the last one (for a general overview of some of these terms, see Nobles 2002). One implication of this difference in word valence is that a measure using one of these terms may tap into something akin to prejudice, while another may merely capture recognition of phenotype. Either way, word choice can shape the inferences one draws through these items.

The translation of survey items also risks generalizing instruments to a population that does not share the same characteristics as the translators of the questionnaire (Ervin and Bower 1952–53). Whether individual translators or focus groups are utilized to vet bilingual instruments, researchers are creating survey items based on the efforts of a few individuals.[4] Here the risk is one of administering survey items that fail to capture concepts as understood by a general population. For instance, Welch, Comer, and Steinman (1973) discovered that Mexican Americans registered differential responses to political questions by language-of-interview.[5] In a more recent analysis of a national survey of Latinos, Lee (2001) finds that respondents interviewed in Spanish systematically registered different opinions than respondents interviewed in English—a finding which remains robust to controls for key political and demographic variables. These studies suggest survey items may not be equally valid across English speakers and Spanish speakers. Yet by examining differences in single survey items, it is difficult to ascertain whether these differences arise from measurement error in individual items, rather than systematic differences in the translated questions. It is in this regard that the current inquiry endeavors to contribute.

What Don't We Know About Item Validity in Bilingual Surveys?

A review of the literature underlines some hazards associated with the development of bilingual survey items. Yet just because things *can go wrong* in the design of bilingual survey items does not mean they

[4]Compare the back translation method of Bobo et al. (1992–94) with that of de la Garza et al. (1989–90).

[5]The authors eventually "eliminate" these language-of-interview effects by controlling for education. But see Lee (2001), who finds that these effects persist in spite of standard demographic and political controls.

do go wrong. Indeed, in translating survey items, it is plausible that measurement error is introduced without systematic bias (e.g., Andrews 1984; Asher 1974; DeShon 1998). For instance, English- and Spanish-speaking Latinos may infer the same meaning from survey items despite some slippage between their English and Spanish wording. In this case, the measures would display high validity (i.e., the items gauge their intended construct across languages). Nevertheless, it is also plausible that while survey questions may, *prima facie*, share a high degree of literal and functional equivalence, in practice the items may fail to elicit uniform meaning across different language groups. These items would be characterized as invalid (i.e., they do not gauge their target construct across languages).

By ignoring the validity of bilingual survey items, scholars risk drawing erroneous conclusions about these items and the traits they seek to measure. As Reise, Widaman, and Pugh explain:

> "To compare groups of individuals with regard to their level on a trait . . . one must assume that the numerical values under consideration are on the same measurement scale . . . If trait scores are not comparable (i.e., on the same measurement scale) across groups, then differences between groups in mean levels or in the pattern of correlations . . . with external variables are potentially artifactual and may be substantively misleading." (1993, 552)

Hence, assessing the invariance of survey items across language groups is not merely a methodological concern. It is also one with profound implications about the inferences one draws when employing survey items to make comparisons between groups of respondents, such as English speakers and Spanish speakers. Indeed, when scholars neglect to verify the cross-language equivalence of measures, the default assumption is that such measures are invariant across different language populations, even though such an assumption may be untenable.

Invariance may not hold for several reasons. For instance, translated items may fail to gauge a concept shared by two or more language groups; or, translated items may tap concepts held by only one linguistic segment of a sampled population. In both instances, respondents will provide answers to the questions being asked by researchers. Yet without further empirical verification, it is unclear whether respondents answer these queries because they tap a preexisting attitude or belief ("real attitudes"), or because respondents wish to meet the demands of the survey interview ("nonattitudes"; e.g., Converse 1964; Tourangeau, Rips, and Rasinski 2000; van der Veld and Saris 2004; Zaller 1992). Distinguishing between "real" and "nonattitudes" in bilingual populations is an important methodological and theoretical concern. For inasmuch as English- and Spanish-speaking Latinos do not share the same concepts, it motivates the development of a stronger grasp of how language shapes one's understanding of political phenomena.

An SEM Approach to Validity in Bilingual Survey Items

To ascertain the validity of bilingual survey items, this paper adopts a structural equation modeling (SEM) strategy. SEM assumes that constructs—such as attitudes, values, or beliefs—are ultimately unobservable. Thus, any given measure will imperfectly tap a concept with a degree of error. The key to valid measurement, then, is to employ multiple indicators. With multiple indicators, SEM parses out systematic variance from measurement error, thus gauging how well survey items capture their intended target.[6]

Let us assume we are interested in a hypothetical political attitude, denoted by ξ. Let us also assume that the researcher has three indicators, or survey questions, to elicit this attitude: x_1, x_2, and x_3. According to SEM, each x, or indicator, gauges attitude ξ with some degree of error. To the extent that these x's covary, SEM corrects for measurement error in each x while representing attitude ξ in terms of $x_1 - x_3$. Thus, the key assumption behind these models is that the covariance between observed variables, $x_1 - x_3$, is explained by an unobserved variable, ξ. In the example at hand, the researcher never "truly" observes attitude ξ; he or she only observes imperfect manifestations of it, as indicated by items $x_1 - x_3$. And so far as these x's covary, one can test whether these x's can be represented as statistical combinations of ξ.

Though SEM is occasionally used by political scientists, this approach is more commonly applied within psychology, where analysts use it to develop and validate measures. One innovation of SEM is that

[6]This is the general case, and it helps to ensure two things: first, that one of these variables is fixed to 1.0 to set the scale of the latent variable; and second, that there will be enough sample variances/covariances to estimate the number of parameters in the model. These conditions are necessary (but not sufficient) to identify a model. For additional background on model identification in an SEM context, see MacCallum (1995) and Kaplan (2000).

it permits one to test whether a set of indicators is invariant—that is, whether it is functionally equivalent—across groups of theoretical interest (e.g., gender, race). Consequently, SEM has enabled scholars to ascertain, not only whether items tap a given construct, but also the degree to which these items tap the same construct across theoretically important groups (Benet-Martinez and John 1998; Byrne, Shavelson, and Muthén 1989; Diaz-Morales et al. 2006; Paunonen and Sampo 1998; Reise, Widaman, and Pugh 1993; Spini 2003). In this regard, SEM can inform researchers as to whether items are fully or partially invariant across groups (e.g., Lambert et al. 2003; Perreira et al. 2005). Full invariance is met when all items attending a concept measure that concept to the same degree across two or more groups. Partial invariance is achieved if some, but not all, items attending a construct measure that construct to the same degree across two or more groups. Though full invariance is the ideal researchers strive for, only partial invariance is needed to compare groups across a given item (Byrne, Shavelson, and Muthén 1989).

In the ideal research world, SEM is an exercise in which strong theory goes hand-in-hand with rich data. By taking this inherently confirmatory approach, researchers are in a position to verify their expectations regarding whether and how far certain items tap a set of traits (Hoyle 1995; Kaplan 2000, 41–52; Kim and Mueller 1978; Kline 2005, 165–206; Saris and Andrews 1991). Many times, however, researchers face a situation where theoretical priors about a set of measures lag behind the availability of data. In these instances, SEM is an exploratory, largely data-driven exercise where the data guides the researcher's decision about which measures tap an attitude or set of attitudes (Hoyle 1995, 2–3; Kaplan 2000, 41–52; Kim and Mueller 1978; Kline 2005, 165–206). Within political science, SEM is an endeavor that falls within these two opposite poles. Many times, political scientists have evolving theories about a particular concept, but the data available to test these theories consist of secondary measures designed by other scholars for their own specific ends. Hence, the typical case is one in which researchers use items from extant surveys (e.g., General Social Survey) to test theories about phenomena they are interested in (e.g., Davis and Brown 2002; de Figueiredo and Elkins 2003; Huddy and Khatib 2007; Hurwitz and Peffley 1987).[7]

I take this third approach in the pending analysis, since it reflects a common situation faced by political scientists. The tradeoff to this approach is that the forthcoming inquiry can only be as good as the availability of data. And, the data we currently have has two limitations that should be noted upfront. First, extant bilingual surveys do not encompass a wide universe of concepts. As in any survey, researchers only ask questions related to particular traits. Ultimately, the number of items addressing these constructs is jointly determined by financial and time constraints: researchers can only ask the number of questions they can afford, without fatiguing respondents. The result is that only a few attitudes are measured with multiple items in any one survey. Moreover, the number and type of bilingual items addressing a particular concept are limited to one-shot surveys, since questions are fielded based on the salience of a construct at the time of a survey. The inferences drawn from the forthcoming analysis will therefore speak less about specific attitudes and beliefs, and more about the general degree of validity in bilingual survey items.

Hypotheses and Data

For each concept examined below, the null hypothesis is that bilingual survey items are equivalent across English- and Spanish-speaking Latinos (H0). This is treated as the null because this assumption is made when survey respondents are pooled across these items in applied work. The logic behind (H0) is that to compare these language groups on a trait—say, American identity—one must assume that both groups are on the same scale. In other words, though two groups differ by language, the statistical relationship between a trait and its indicators is presumed to be equivalent across both groups, chance variations aside.

Two hypotheses are tested against this null. The first hypothesis, (H1), anticipates that net of measurement error, the specified items fail to collectively measure the same construct across English-speaking and Spanish-speaking Latinos. In other words, this hypothesis expects that linguistic differences between items compromise their cross-language validity. The second hypothesis (H2) builds on the first by assessing the extent to which the cross-language validity of items is compromised by some, rather than all, indicators in the measurement models.

[7]For examples from political science, psychology, and sociology, respectively, see Citrin, Reingold, and Green (1990), Chan et al. (2007), and Perreira et al. (2005).

These hypotheses are assessed on five constructs: (1) American identity; (2) black stereotypes; (3) spending preferences; (4) immigration preferences; and (5) perceptions of discrimination. Two considerations drove selection of these concepts. First, each had to be gauged by enough items to identify and estimate a measurement model. Second, each has been or can be used to analyze Latino public opinion. This means the data must test a construct that arises from a body of literature which *has* or *can be* extended to Latino public opinion. This ensures a body of theoretical work to guide our expectations about a concept. In the interest of space, I use footnotes to refer readers to pertinent literature on each construct. These references are meant to be exemplary and suggestive, but by no means exhaustive.

The first construct I examine is American identity, as measured in the *Pew National Survey of Latinos (2004)*.[8] The items for this analysis consist of four ''agree-disagree'' questions about specific traits that an immigrant must hold in order to be considered part of American society: (1) formal citizenship; (2) belief in the U.S. Constitution; (3) inclination to vote; and (4) ability to speak English. This model thus assesses whether and to what degree these four items gauge notions of American identity across English- and Spanish-speaking Latinos.

The second construct I examine is black stereotypes.[9] Here I test the validity of seven (7) items gauging belief in stereotypical traits about blacks, as found in the *Multi-City Study of Urban Inequality (1992–94)*. These traits are blacks: (1) as unintelligent; (2) as poor; (3) as welfare-dependent; (4) as hard to get along with; (5) as poor English speakers; (6) as involved with drugs and gangs; and (7) as discriminatory toward others.[10]

The third concept I consider is spending preferences. This analysis assesses whether traditional notions of ideology bind spending preferences across issue domains, such that greater liberalism leads to increased support for spending across issues.[11] Thus, I estimate a model of Latino spending preferences using 10 categorical items from the *Latino National Political Survey (1989–90)*. These items correspond to 10 domains: (1) the environment; (2) public education; (3) welfare; (4) health; (5) science; (6) children's services; (7) refugees; (8) defense; (9) crime; and (10) blacks. Respondents were invited to register a preference to decrease spending in each program, leave spending in each program the same, or increase spending in each program.

Next, I consider the degree to which various items tap a general preference for immigration.[12] This model employs data from the *Kaiser/Pew Latino Survey (2002)*. Using two trichotomous and two dichotomous items that address immigration, I test whether they are invariant across English- and Spanish-speaking Latinos. The trichotomous items gauge preferences toward (1) the overall level of immigration, and (2) legal immigration from Latin America. Respondents registered support for less, the same amount, or more immigration. The dichotomous items gauged preferences for (3) amnesty for illegal immigrants, and (4) a guest worker program. Here

[8]For historical background on various conceptualizations of American identity, see Higham (1955); Jacobson (1998); and King (2000). For insight on the empirical measurement of American identity (and its variant manifestations) as well as its implications for political behavior, see Kosterman and Feshbach (1989); Citrin, Reingold, and Green (1990); Citrin, Wong, and Duff (2001); de Figueiredo and Elkins (2003); and Huddy and Khatib (2007). For background on the relationship between Latinos and some of the components of American identity, see de la Garza, Falcon, and Garcia (1996); Hood, Morris, and Shirkey (1997); and Citrin et al. (2007).

[9]On measures of black stereotypes and their political effects, see Gilens (1996, 1999); Sears et al. (1997); and Bobo and Johnson (2000). For historical perspective on immigrant communities' belief in black stereotypes, see Shankman (1982); Hellwig (1982); and Ignatiev (1995). For research on Latinos and black stereotypes, see McClain et al. (2006).

[10]The items originally ran from 1 to 7. However, the proportion of responses in some answer categories was so low that they yielded insufficient information to estimate the multiple thresholds of the items. I therefore collapsed the original categories into two, 0 and 1, in order to reduce the number of thresholds to be estimated, and thus, enable the estimation of a model without substantial loss of information. Values between 1 through 4 were recoded as 0 (''no belief in stereotypical trait''), while values above 4 were recoded as 1 (''belief in stereotypical trait'').

[11]For general background on the traditional conceptualization of ideology and its effects on public opinion, see Converse (1964), Achen (1974), Stimson (1991), and Zaller (1992). For insight into the structure and influence of political ideologies among blacks and Latinos, see Dawson (2001) and Uhlaner and Garcia (2001), respectively.

[12]For theoretical background on models of anti-immigrant opinion, see Tolbert and Hero (1996); Citrin et al. (1997); Fetzer (2000); Alvarez and Butterfield (2000); and Kessler (2001). For insight into Latino attitudes toward immigration—and their similarities and dissimilarities to white Americans—see Miller, Polinard, and Wrinkle (1984); de la Garza et al. (1991); Hood, Morris, and Shirkey (1997); Hood and Morris (1997).

respondents indicated support or opposition to each item.[13]

Finally, I consider perceptions of discrimination. This model assesses whether personal views of discrimination are functionally equivalent across English-speaking and Spanish-speaking Latinos by using seven items from the *Latino National Politics Survey (1989–90)*.[14] The items tap perceptions of discrimination toward: African Americans, Mexicans, Cubans, Puerto Ricans, Women, and Jews. Respondents registered whether each group experiences "no discrimination," "a little discrimination," "some discrimination," or "a lot of discrimination".

Model and Patterns of Evidence

The general model for the pending analyses is such that each indicator x_m, where $m = 1, \ldots n$, is configured as a function of the unobserved political construct, ξ, plus a random error term. This relationship is represented in the following form,

$$x_m = \lambda_m \xi + \delta_m \tag{1}$$

where λ_m is the estimate of the regression of x_m on ξ. Since each construct is attended by four or more observed indicators, the above model is distilled into the following matrix form,

$$X = \Lambda \xi + \delta \tag{2}$$

where X is a $(n \times 1)$ column vector of scores of respondent i on n survey indicator, Λ is a matrix of loadings of the n indicator on ξ, and δ is the matrix of measurement residuals.[15]

Hypothesis testing for each concept adheres to the following sequence. First, one must estimate a well-fitting baseline model for both language groups. Consider American identity. Here the baseline model places no restrictions on the loadings and thresholds of the items tapping this trait. This means the relationship between each item and American identity is freely estimated across both language groups. The objective here is to produce a model with excellent fit. At minimum, this entails a model where the item loadings are positive and statistically significant.[16] Moreover, it involves assessing a model's fit through conventional indices, in this case: CFI (Comparative Fit Index), TLI (Tucker-Lewis Index), and RMSEA (Root Mean Square Error of Approximation). Models with a CFI and TLI value above .90, and an RMSEA value below .10, will be deemed as having a good fit.[17] A well-fitting model will then serve to evaluate subsequent models, which test our hypotheses by fixing to equality the loadings and thresholds of the individual items across both language groups.

Once a well-fitting baseline model is obtained, the attendant items are fixed to equality across English-speaking and Spanish-speaking Latinos. Evidence against the null of full invariance will consist of a statistically significant change in the chi-square (χ^2) of the resulting model. Or, to put it differently, one will reject this null hypothesis if these equality constraints across both language groups lead to a statistically significant change in chi-square. The test for partial invariance is a slight modification of the test above. Rather than fixing all items to equality across both language groups, one freely estimates one item while constraining the rest to equality. And inasmuch as

[13]It is plausible that preferences for immigration depend on whether the policies target illegal or legal immigration. But alas, there were insufficient items tapping both varieties of policy to estimate a two-factor model to address this possibility. Future research should investigate this as richer data become available.

[14]Perceptions of discrimination shape group-based identities among racial minorities, as evidenced by research on African-American political behavior (e.g., Dawson 1994; Shingles 1981; Tate 1994). Following this lead, Latino politics scholars have discovered that perceptions of discrimination are consequential for Latino political behavior (e.g., de la Garza, Garcia, and Falcon 1992; Garcia 2003; Masuoka 2006; Sanchez 2006; Uhlaner and Garcia 2002).

[15]In the general case, equation (1) is estimated using a sample covariance matrix based on normally distributed variables. Given the categorical nature of our observed indicators, the data at hand is inherently nonnormal. This requires two modifications: the inclusion of a threshold model and the use of polychoric correlations as input. These features and their relationship to the general model are explained in further detail in the online Appendix D.

[16]Items are deemed statistically significant at the 5% level or better. In most cases, statistically insignificant items are dropped from further analysis if their omission improves the fit of the baseline model. This is because the invariance tests that follow are premised on a series of nested models: one where the parameters for the indicators are freely estimated across both groups, and the other where these parameters are fixed to equality across both groups. It is therefore imperative that the unrestricted model achieves as good a fit as possible, since the restricted model will be evaluated against it. Simply put, it is pointless to compare two nested models if the initial one is of dubious fit.

[17]These thresholds follow those generally recommended by structural equation modelers (Browne and Cudeck 1993; Kline 2005, 133–45). SEM practitioners suggest assessing model fit through more than one fit index, since any one index gauges particular aspects of model fit. For instance, CFI and TLI are both incremental indices, such that they compare the fit of a researcher's model to a baseline model where zero covariances are assumed. In some cases, then, researchers may obtain a well-fitting model that is not necessarily parsimonious. In contrast, RMSEA is a parsimony-adjusted index that favors simpler models. RMSEA reflects the degree to which a researcher's model fits the population covariance matrix, while taking the degrees of freedom and sample size into account.

TABLE 1 American Identity: Parameter Estimates and Fit Statistics for Baseline, Full Invariance, and Partial Invariance Models

Indicators: Loadings & Thresholds	Baseline Model (1): English-Speakers	Baseline Model (1): Spanish-Speakers	Full Invariance Model	Partial Invariance Model (1)	Partial Invariance Model (2)	Partial Invariance Model (3)
Speaks English	1.00	1.00	1.00	1.00	1.00	1.00
τ_{11}	−.13	−.18				
Believes in Constitution	1.04	.92	Fixed	Free	Fixed	Fixed
	(.20)	(.11)				
τ_{21}	−.87	−1.03				
Is a U.S. Citizen	1.27	1.19	Fixed	Fixed	Free	Fixed
	(.21)	(.11)				
τ_{31}	−.16	−.15				
Votes in Elections	1.39	1.20	Fixed	Fixed	Fixed	Free
	(.24)	(.11)				
τ_{41}	−.16	−.69				
TLI	.96	—	.95	.96	.93	.98
CFI	.98	—	.96	.97	.96	.99
RMSEA	.05	—	.06	.05	.06	.04
χ^2	14.99	—	35.80	23.04	32.19	14.19
χ^2 change	—	—	20.34	7.99	16.03	.31†

Notes: For all equations, N = 2,278. All parameter estimates and changes in chi-square in all models are significant at the 5% level or better, unless otherwise indicated by †. Standard errors are in parentheses. All data are weighted. First item loading in each model is fixed to 1.00 for identification.

these single tests yield a statistically significant change in chi-square, one has evidence that the freely estimated indicator is not invariant across both groups.

All models are estimated in Mplus® using robust WLS with polychoric correlations.[18] Given the discrete nature of the data being used, this procedure avoids violating the assumption of multivariate normality. Throughout, I report the unstandardized coefficients along with their standard errors as a way for the reader to gauge the statistical reliability of the estimates.

Results

I begin by examining the results for American identity. This concept is measured using four binary items which gauge respondent agreement with whether an immigrant must hold each of the following traits in order to be part of American society: (1) be a U.S. citizen; (2) speak English; 3) believe in the Constitution; and (4) vote in U.S. elections. To that end, I estimated a baseline model where all four of these indicators are freely estimated.

Looking at Table 1, two things stand out. First, the baseline model displays excellent fit, as noted by our CFI, TLI, and RMSEA, which surpass the benchmarks set out earlier. Moreover, the loadings for each item are statistically significant at the 5% level or better. Given the excellent fit of this baseline model, I fix to equality the loadings and thresholds of all four items across both language groups. These constraints produce a statistically significant change in chi-square, thus allowing one to reject the null hypothesis of full invariance. In other words, fixing these parameters to equality across both language groups of Latinos results in a poorer model fit than when the same parameters are freely estimated, which suggests that these items are not equivalent across linguistic groups. I then test for partial invariance by freely estimating the loadings and thresholds of each indicator while fixing the rest to equality. This set of tests reveals that only one item, *votes in elections*, is invariant across both groups, as evidenced by the statistically insignificant change in chi-square. Hence,

[18]Robust WLS (Weighted Least Squares) is the estimator recommended by modelers when data are nonnormal and categorical, as is the case in the analysis at hand (see Finney and DiStefano 2006). In using this estimator, a corrected chi-square difference test is employed. This corrected statistic is needed since the difference in chi-square values between two nested models is not distributed as a chi-square when using robust WLS. Thus, the quantity of interest in testing the invariance of items is not the exact difference between the chi-squares of the nested models (which is adjusted), but rather, whether the more restrictive of the nested models yields a deterioration in model fit, as captured by statistically significant change in the corrected chi-square.

with the exception of this single item, these survey questions are not comparable across English- and Spanish-speaking Latinos.

Next, I turn to the findings for black stereotypes. I begin by examining the results from an initial baseline model for this construct. Judging by the fit statistics, this initial baseline model achieves a poor level of fit, as both the TLI and CFI statistic fall well below the .90 target (.71 and .75, respectively). Further inspection reveals that three of the items—*poor*, *speaks poor English*, and *discriminatory toward others*—are statistically insignificant: a pattern which comports with the fit indices for this model. I therefore estimated a second baseline model without these three items. Looking at the fit indices for this follow-up exercise, one sees that though some of the fit indices improve incrementally, they generally fail to reach the targets set earlier.[19] Moreover, additional items—such as *involved with drugs & gangs* and *prefers welfare*—yield statistically insignificant estimates. Thus, while no formal invariance tests were possible with this data, the two attempts at a baseline model suggest that the items are laden with measurement error, to the point that they fail to systematically tap the trait at hand.

For spending preferences, I again started with a basic model where all items are assumed to measure the same construct. As Table 3 notes, our fit statistics indicate an excellent fit for this initial model. Yet inspection of the loadings indicates that the item *defense spending* is statistically insignificant among English speakers. I thus estimated a new baseline model without this item. Having done this, the model fit improves incrementally: each fit statistic surpasses the anticipated benchmarks, while the remaining nine items retain statistical significance.

Based on this last model, I constrained to equality the loadings and thresholds for all of the indicators across both language groups, which yielded a statistically significant change in chi-square. This indicates the parameter constraints deteriorate the model's fit, thereby reducing one's confidence about the linguistic equivalence of these items. I then freely estimated each indicator on a one-by-one basis, while constraining the remaining ones to equality. And in each instance, this test yielded a statistically significant change in chi-square. This evidence thus suggests that collectively and in isolation, these items are not linguistically equivalent.

To gauge immigration preferences, I assessed how well four items tapped this concept among Latinos of different language backgrounds. I began by fitting a baseline model where all four items are presumed to measure this construct. This model achieved an excellent fit, as evidenced by the fit statistics in Table 4 (TLI and CFI $> .90$; RMSEA $< .05$). However, the item *guest worker* yields a statistically insignificant estimate for English and Spanish speakers. Given the indications of a well-fitting model in spite of this one insignificant estimate, I proceeded to run tests for full and partial invariance.

If one constrains to equality the loadings and thresholds across both language groups, a statistically significant change in chi-square is yielded. If one freely estimates the loading and thresholds for each indicator, while fixing the rest of these parameters to equality, a similar pattern emerges: a statistically significant change in chi-square. Thus, either collectively or in isolation, these items do not appear to be statistically equivalent across language groups.[20]

Finally, when we turn to perceptions of discrimination, the initial baseline model generally displays indications of a good fit. As seen in Table 5, the CFI and TLI for this baseline model meet or exceed the .90 standard set out. Moreover, the RMSEA for this model falls at .10, and the parameter estimates for each of the items are statistically significant at the 5% level.

Given these model characteristics, I test for full and partial invariance, respectively. The results in Table 5 suggest that invariance does not characterize these items, since imposing equality constraints on the items results in a statistically significant chi-square change. Thus, collectively these indicators are not gauging the same trait across both English- and Spanish-speaking Latinos. I then tested for partial invariance by freely estimating each indicator while fixing the rest to equality across both language groups. And as Table 5 reveals, each of these tests yields a statistically significant increase in chi-square. Taken together, this evidence suggests that these items are not interchangeable across English- and Spanish-speaking Latinos.

[19]One may reasonably contend that the fit indices of these alternative unrestricted models are not strictly comparable, since dropping statistically insignificant indicators produces models based on slightly different covariance-variance matrices. Yet I consider this a more conservative strategy than comparing two nested models when the unrestricted model is of dubious fit. Indeed, in establishing an unrestricted model with reasonably good fit, one is making it more difficult to reject this model through subsequent invariance testing, which place equality constraints on the key parameters of interest.

[20]It is plausible that these results arise from one item's (*guest worker*) lack of statistical significance. One strategy is to fix this item's loading to zero across both language groups and then rerun the analysis. Doing so leads to the same substantive result: a lack of invariance across the remaining items. These results can be found in online Appendix B.

Table 2 Black Stereotypes: Parameter Estimates and Fit Statistics for Baseline Models

Indicators: Loadings & Thresholds	Baseline Model (1): English-Speakers	Baseline Model (1): Spanish-Speakers	Baseline Model (2): English-Speakers	Baseline Model (2): Spanish-Speakers
Unintelligent	1.00	1.00	1.00	1.00
τ_{11}	.69	.49	.69	.49
Involved w/drugs & gangs	1.00	1.35	.52†	1.47†
	(.38)	(.64)	(.31)	(.79)
τ_{21}	−.51	−.67	−.51	−.67
Poor	.75†	.86†	—	—
	(.47)	(.53)		
τ_{31}	−.49	−.57		
Prefers welfare	.85	1.46	.35†	1.51†
	(.35)	(.72)	(.24)	(.92)
τ_{41}	−.51	−.96	−.51	−.96
Hard to get along with	1.04	1.17	.67	1.14†
	(.31)	(.60)	(.33)	(.64)
τ_{51}	.59	−.15	.59	−.15
Speaks poor English	.30†	.80†	—	—
	(.36)	(.63)		
τ_{61}	.64	.95		
Discriminatory to others	.53†	1.32	—	—
	(.34)	(.66)		
τ_{71}	−.54	−.30		
TLI	.71	—	.75	—
CFI	.75	—	.89	—
RMSEA	.05	—	.06	—
χ^2	28.05	—	8.14	—

Notes: For all equations, N = 573. All parameter estimates significant at the 5% level or better, unless otherwise indicated by †. Standard errors are in parentheses. All data are weighted. First item loading in each model is fixed to 1.00 for identification.

The Consequences of Lacking Linguistic Invariance

The lack of linguistic invariance suggests bias in the measurement of political attitudes across English- and Spanish-speaking Latinos. But what is the direction and magnitude of this bias? To answer this, I calculated the average proportions of English speakers and Spanish speakers classified into key response categories by the attendant items. Consider the trait American identity, where agreement with its attendant indicators is theorized to reveal a belief in this concept. Here I took the average proportion of each language group indicating agreement across all four items gauging this concept. I then assessed the degree of difference in these proportions and interpreted this quantity as one sign of the magnitude and direction of the bias.

This approach yielded statistically significant differences between both language groups across those traits where linguistic invariance was not met. And as Table 6 below reveals, the magnitude of bias ranges anywhere from 6% to 10%, according to this approach.

One further sees that the direction of bias is inconsistent: Sometimes English speakers are classified as displaying higher levels of a trait, while other times, Spanish speakers play this role. All of this is to say that the bias induced by the lack of linguistic invariance is not negligible, especially since the group affected by the bias changes on a construct-by-construct basis.

What about Class and Cognitive Sophistication?

To this point, the evidence has generally suggested that the survey items under investigation are not

Table 3 Spending Preferences: Parameter Estimates and Fit Statistics for Baseline and Full Invariance Models

Indicators: Loadings & Thresholds	Baseline Model (1): English-Speakers	Baseline Model (1): Spanish-Speakers	Baseline Model (2): English-Speakers	Baseline Model (2): Spanish-Speakers	Full Invariance Model	Partial Invariance Model(1)	Partial Invariance Model(2)
Environment	1.00	1.00	1.00	1.00	1.00	1.00	1.00
τ_{11}	−1.96	−1.60	−1.96	−1.60			
τ_{12}	−.56	−.41	−.56	−.41			
Education	1.58	1.27	1.59	1.27	Fixed	Free	Fixed
	(.22)	(.12)	(.22)	(.12)			
τ_{21}	−2.28	−2.21	−.228	−2.21			
τ_{22}	−1.06	−1.02	−1.06	−1.02			
Welfare	.84	.73	.83	.72	Fixed	Fixed	Free
	(.17)	(.10)	(.17)	(.10)			
τ_{31}	−.74	−1.02	−.74	−1.02			
τ_{32}	.28	.01	.28	.01			
Health	1.21	1.22	1.22	1.22	Fixed	Fixed	Fixed
	(.17)	(.12)	(.18)	(.12)			
τ_{41}	−1.95	−2.33	−1.95	−2.33			
τ_{42}	−.76	−.98	−.76	−.98			
Science	.89	1.18	.86	1.15	Fixed	Fixed	Fixed
	(.14)	(.11)	(.14)	(.10)			
τ_{51}	−1.23	−1.64	−1.23	−1.64			
τ_{52}	.14	−.38	.14	−.38			
Child Services	1.19	1.36	1.19	1.35	Fixed	Fixed	Fixed
	(.18)	(.12)	(.17)	(.11)			
τ_{61}	−1.73	−2.31	−1.73	−2.31			
τ_{62}	−.49	−.88	−.49	−.88			
Refugees	1.15	1.20	1.16	1.19	Fixed	Fixed	Fixed
	(.19)	(.10)	(.19)	(.09)			
τ_{71}	−1.15	1.20	1.16	1.81			
τ_{72}	.15	−.79	.15	−.79			
Defense	.17[†]	.64	—	—	—	—	—
	(.13)	(.08)					
τ_{81}	−.18	−.74					
τ_{82}	.95	.23					
Crime	1.30	.91	1.29	.90	Fixed	Fixed	Fixed
	(.21)	(.11)	(.21)	(.11)			
τ_{91}	−1.95	−1.69	−1.95	−1.69			
τ_{92}	−1.21	−1.31	−1.21	−1.31			
Blacks	1.25	1.16	1.25	1.13	Fixed	Fixed	Fixed
	(.20)	(.10)	(.20)	(.10)			
τ_{101}	−1.56	−1.67	−1.56	−1.67			
τ_{102}	−.17	−52	−.17	−.52			
TLI	.93	—	.94	—	.93	.83	.89
CFI	.93	—	.94	—	.92	.79	.87
RMSEA	.05	—	0.5	—	.05	.07	.06
χ^2	185.00	—	150.83	—	197.01	442.76	288.53
χ^2 Change	—	—	—	—	66.08	260.07	142.98

Table 3 (Continued)

Indicators: Loadings & Thresholds	**Partial Invariance Model(3)**	**Partial Invariance Model(4)**	**Partial Invariance Model(5)**	**Partial Invariance Model(6)**	**Partial Invariance Model(7)**	**Partial Invariance Model(8)**
Environment	1.00	1.00	1.00	1.00	1.00	1.00
τ_{11}						
τ_{12}						
Education	Fixed	Fixed	Fixed	Fixed	Fixed	Fixed
τ_{21}						
τ_{22}						
Welfare	Fixed	Fixed	Fixed	Fixed	Fixed	Fixed
τ_{31}						
τ_{32}						
Health	Free	Fixed	Fixed	Fixed	Fixed	Fixed
τ_{41}						
τ_{42}						
Science	Fixed	Free	Fixed	Fixed	Fixed	Fixed
τ_{51}						
τ_{52}						
Child Services	Fixed	Fixed	Free	Fixed	Fixed	Fixed
τ_{61}						
τ_{62}						
Refugees	Fixed	Fixed	Fixed	Free	Fixed	Fixed
τ_{71}						
τ_{72}						
Defense	—	—	—	—	—	—
τ_{81}						
τ_{82}						
Crime	Fixed	Fixed	Fixed	Fixed	Free	Fixed
τ_{91}						
τ_{92}						
Blacks	Fixed	Fixed	Fixed	Fixed	Fixed	Free
τ_{101}						
τ_{102}						
TLI	.86	.90	.89	.93	.89	.89
CFI	.88	.88	.87	.91	.86	.87
RMSEA	.07	.06	.06	.05	.06	0.6
χ^2	300.33	259.73	289.03	208.72	295.50	286.85
χ^2 Change	150.96	121.62	145.58	72.77	146.26	140.73

Notes: For all equations, N = 2,636. All parameter estimates and changes in chi-square are all significant at the 1% level, unless otherwise indicated by †. Standard errors are in parentheses. All data are weighted. First item loading in each model is fixed to 1.00 for identification.

invariant across language. Yet these results can be plausibly explained, not by differences in language, but by differences in income and education.[21] In other words, the observed lack of linguistic invariance stems from class differences (as indexed by income) and/or differences in cognitive sophistication (as indexed by education). But so far as income and education, rather than language, are responsible for the lack of item equivalence in these survey items, what general pattern in the data should we expect to unearth? I propose that the clearest evidence would be the following: once one accounts for differences in income and education, the items should reach

[21]I thank one of the anonymous reviewers for suggesting this explanation, as well as a strong empirical test of it.

TABLE 4 **Immigration Policy Preferences: Parameter Estimates and Fit Statistics for Baseline and Full and Partial Invariance Models**

Indicators: Loadings & Thresholds	Baseline Model (1): English-Speakers	Baseline Model (1): Spanish-Speakers	Full Invariance Model	Partial Invariance Model (1)	Partial Invariance Model (2)	Partial Invariance Model (3)
Level of Immigration	1.00	1.00	1.00	1.00	1.00	1.00
τ_{11}	−1.45	−1.36				
τ_{12}	.04	−.14				
Legal Immigration	1.23	2.84	Fixed	Free	Fixed	Fixed
	(.18)	(1.30)				
τ_{21}	−.33	.15				
τ_{22}	.75	1.21				
Amnesty for Illegals	1.18	2.16	Fixed	Fixed	Free	Fixed
	(.14)	(.56)				
τ_{31}	.96	2.12				
Guest Worker Program	.08†	.13†	Fixed	Fixed	Fixed	Free
	(.10)	(.25)				
τ_{41}	.38	.13				
TLI	.94	—	.80	.97	.69	.67
CFI	.98	—	.86	.95	.79	.77
RMSEA	.03	—	.06	.03	.08	.08
χ^2	10.77	—	48.11	15.80	71.31	76.28
χ^2 change	—	—	41.19	7.75	64.40	71.81

Notes: For all equations, N = 2,923. All parameter estimates and changes in chi-square are all significant at the 1% level, unless otherwise indicated by †. Standard errors are in parentheses. All data are weighted. First item loading in each model is fixed to 1.00 for identification.

full invariance across English speakers and Spanish speakers.

I thus classified respondents for each construct into high and low categories of education and income by using a median split. This strategy yielded four (4) subsamples of respondents for each concept: (1) low-income respondents; (2) high-income respondents; (3) low-education respondents; and (4) high-education respondents. Within each of these subsamples, the linguistic invariance of items was tested for American identity; perceptions of discrimination; and spending preferences, thereby yielding 12 tests of the alternate explanation at hand.[22]

Recall the pattern we expect to uncover via these tests: the items attending each trait should reach invariance across language when we hold constant income and education differences. Yet the results from the 12 invariance tests, reported below in Table 7, generally contradict this expectation. In 9 of the 12 tests, holding income or education constant did not produce linguistically invariant items.

Of the three tests that yielded linguistically invariant items, the results were produced only when levels of education were held constant. And of these three, two were yielded among low-education respondents, and one among high-education respondents. The predominant pattern, therefore, is one where the lack of linguistic invariance persists, even after considering these plausible alternative influences. While not completely ruling out these alternate explanations, the evidence does increase one's confidence that language is not a spurious explanation for the lack of invariance displayed by these items.

What Have We Learned? Discussion and Recommendations

Researchers generally presume that bilingual survey items are comparable across language groups. The preceding analyses, however, revealed no consistent

[22]One construct—immigration preferences—was not amenable to further analysis. The transformation of the data into the requisite subsamples described above yielded relatively low proportions of responses in some item categories, such that the ability of any subsequent model to converge was hindered.

TABLE 5 Perceived Discrimination: Parameter Estimates and Fit Statistics for Baseline Model, Full Invariance Model, and Partial Invariance Models

Indicators: Loadings & Thresholds	Baseline Model: English-Speakers	Baseline Model: Spanish-Speakers	Full Invariance Model	Partial Invariance Model(1)	Partial Invariance Model(2)	Partial Invariance Model(3)	Partial Invariance Model(4)	Partial Invariance Model(5)	Partial Invariance Model(6)
African Am.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
τ_{11}	−1.92	−1.08							
τ_{12}	−1.21	−.42							
τ_{13}	.04	.45							
Asian Am.	.76	.97	Fixed	Free	Fixed	Fixed	Fixed	Fixed	Fixed
	(.07)	(.04)							
τ_{21}	−1.21	−.38							
τ_{22}	−.36	.53							
τ_{23}	.99	1.56							
Mexicans	1.14	1.03	Fixed	Fixed	Free	Fixed	Fixed	Fixed	Fixed
	(.06)	(.04)							
τ_{31}	−1.71	−1.16							
τ_{32}	−.91	−.34							
τ_{33}	.50	.48							
Cubans.	1.31	1.11	Fixed	Fixed	Fixed	Free	Fixed	Fixed	Fixed
	(.07)	(.04)							
τ_{41}	−1.58	−.65							
τ_{42}	−.61	.20							
τ_{43}	.76	1.20							
Puerto Ric.	1.32	1.12	Fixed	Fixed	Fixed	Fixed	Free	Fixed	Fixed
	(.07)	(.03)							
τ_{51}	−1.44	−.49							
τ_{52}	−.66	.31							
τ_{53}	.82	1.18							
Women	.87	1.04	Fixed	Fixed	Fixed	Fixed	Fixed	Free	Fixed
	(.06)	(.04)							
τ_{61}	−1.06	−.37							
τ_{62}	−.37	.33							
τ_{63}	.66	1.06							
Jewish Am.	.90	.88	Fixed	Fixed	Fixed	Fixed	Fixed	Fixed	Free
	(.07)	(.05)							
τ_{17}	−.75	−.09							
τ_{72}	−.06	.58							
τ_{73}	1.07	1.39							
TLI	.96	—	.97	.88	.86	.88	.88	.87	.87
CFI	.94	—	.93	.83	.78	.81	.82	.80	.81
RMSEA	.10	—	.09	.16	.18	.17	.16	.17	.17
χ^2	270.69	—	348.68	804.06	1000.90	860.84	825.79	900.07	873.84
χ^2 change	—	—	126.86	454.82	614.72	515.72	491.69	548.36	498.34

Note: For all equations, N = 2,642. All parameter estimates and changes in chi-square are all significant at the 1% level, unless otherwise indicated by †. Standard errors are in parentheses. All data are weighted. First item loading in each model is fixed to 1.00 for identification.

TABLE 6 Language Gap by Political Construct

	American Identity	Immigration Preferences	Discrimination Perceptions	Spending Preferences
English speakers	62% (1068)	31% (1262)	26% (1090)	64% (1089)
Spanish speakers	68% (1210)	23% (1661)	17% (1552)	74% (1547)
Language Gap	6%	8%	9%	10%

Notes: The number of cases for each language group is in parentheses. Differences in percentages are statistically significant at the 5% level or better. Response categories assessed for items attending each construct are as follows: American Identity (agrees with American trait); Immigration Preferences (supports restrictions); Discrimination Perceptions (perceives a lot of discrimination); and Spending Preferences (supports increases in spending).

pattern to this effect. Out of five separate analyses on constructs from four different surveys, the null of invariance across English-speaking and Spanish-speaking Latinos was rejected in four instances. The exceptions to this pattern were the items gauging black stereotypes. Here the statistical unreliability of these estimates was such that it hindered the estimation of a baseline model that could subsequently be tested in a multigroup framework—an indication that the items are inherently laden with error, and perhaps capturing "nonattitudes" (e.g., Converse 1964). Moreover, when examining each concept on an item by item basis, it was discovered that only a single item was invariant across language groups—*votes in U.S. elections*, in the American identity analysis. What, then, can we learn from these findings? I believe the lessons and implications of these findings can be distilled into two varieties, those dealing with applied analyses using extant data, the other touching on the development of future bilingual items.

Taken as a whole, this investigation reveals that insofar as our tests and constructs are concerned, item validity is weak in U.S. bilingual political surveys. In this regard, this research extends the research of Lee (2001) and others (e.g., Welch, Comer, and Steinman 1973) by yielding evidence which suggests that language differences in survey responses exist independent of measurement error and across multiple surveys of Latinos. Indeed, these language differences generally persisted even after holding constant levels of income and education, respectively. These latter findings increase our confidence that the observed language differences are not spurious, while emphasizing that different language groups may sometimes not uniformly share political traits.

Nevertheless, the evidence in this paper does not mean that extant data cannot continue to facilitate research on Latino public opinion. It does mean, however, that a key assumption in current research designs may sometimes be untenable. When pooling observations across language, researchers presume that respondents are all on the same scale of a given attitude. Yet this is an assumption which can be empirically verified. And if the assumption does not fit, one strategy around this quandary is to disaggregate one's analyses by language, rather than pooling observations across language. This ensures that extant bilingual data can be marshaled toward new analyses without violating the critical assumption of items' linguistic invariance.

Still, the above recommendation speaks only to what researchers can do to work with available bilingual survey data. It does nothing to address the more important question of what scholars can do in the future in light of the findings of this analysis. One important step in this regard is to more thoroughly validate survey items prior to administration. It is not that current methods (i.e., back translation) are

TABLE 7 Invariance Test Results When Holding Constant Socioeconomic Status

	High Income	Low Income	High Education	Low Education
American Identity	No	No	Yes	Yes
Spending Preferences	No	No	No	Yes
Perceived Discrimination	No	No	No	Yes

Notes: Yes means that invariance was achieved. No means invariance was not met. These results refer to those yielded through full invariance tests. Partial invariance tests generally parallel these findings. To preserve space, the full set of results is reported in web appendix I.

incorrect, so much as they are incomplete. One of the virtues of using standard back-translation methods is that they are a cost-effective way to design survey items. But the additional steps needed to enhance the validity of bilingual survey items need not add excessively to the already exorbitant cost of bilingual surveys. For instance, scholars can test translated items on local samples of convenience that are matched on characteristics deemed important (e.g., language, education, etc.). To optimize the rigor of these tests, scholars can embed simple wording experiments to assess whether certain word choices in translated items lead to statistically significant changes in responses (McDermott 2002; Shadish, Cook, and Campbell 2002). Scholars might also develop future bilingual survey items using techniques which assess and correct for the incomparability of response categories (e.g., King et al. 2004). In short, researchers have at their disposal an array of tools to complement their back-translation efforts.

Though it is easy to interpret the findings in this study as a rebuke on bilingual survey instruments, the evidence is actually a sign of promise for survey researchers in general, and scholars of Latino public opinion in particular. That these bilingual survey items did not perform up to expectations suggests the immediate need for greater theorizing about concepts and the items used to measure them. The fact that the items in this paper failed to achieve invariance across language groups means just that: that *these* items are not invariant. This evidence, though, should not be construed as implying that these traits do not exist in our population of interest. It only implies that the items at hand did not uniformly capture their target constructs.

For scholars of Latino public opinion, then, the challenge ahead of us is to develop stronger theories regarding the concepts we employ in our research designs. By necessity, this entails new research that yields a more nuanced understanding about what type of concepts are important to which Latinos, what the structure of these constructs is, what type of items best gauge these phenomena, and whether and why these indicators (should) extend across English- and Spanish-speaking Latinos. Such efforts do not have to await the arrival of new omnibus Latino surveys. The intellectual creativity of scholars, combined with a mix of methods (e.g., focus groups and experimental designs) can be harnessed to provide deeper insight into the structure and measurement of Latino public opinion. Consider the items tapping American identity. Scholars can build on the current findings by designing and administering a richer battery of items that tap into related yet distinct notions of national identity, such as nationalism and patriotism. This exercise would inject greater nuance into our understanding about the manifestations of national attachments among Latinos and produce a deeper understanding of the structure of these distinct attachments across this linguistically varied group.[23]

Of course, inasmuch as the findings of this paper may advance our understanding of bilingual survey data, the evidence still rests on a specific research design—in this case, one that was preoccupied with the role of measurement error in bilingual survey items. Thus, additional analyses using different techniques would be a welcome and useful step to refine our understanding about the intersection between language and survey response. For instance, in addition to the role of measurement error, scholars might be concerned with the discriminating power of individual survey items. That is, even if survey items are tapping the same construct, are these items yielding the same statistical information across language groups (Lambert et al. 2003)? To that end, future work may employ Item Response Theory (IRT) on bilingual data to ascertain the extent to which items provide similar information across English-speaking and Spanish-speaking Latinos (e.g., Treier and Jackman 2008).[24] Research along these lines would deepen our understanding about the psychometric properties of bilingual survey data.

Future research should also shed further light on the degree to which the lack of invariance stems from self-selection processes that generally affect all survey data.[25] This paper discovered that once respondents choose to complete surveys in one language over the other, we observe a lack of invariance in some survey items. Future research can deepen our knowledge here by exercising stronger control over the choice of language-of-interview. In particular, one can imagine conducting an experiment among individuals who are equally comfortable in their ability to speak English and Spanish, and then randomly

[23]This suggestion elaborates on a point constructively suggested by an anonymous reviewer. For more on the varieties of national attachments, see Huddy and Khatib (2007) and Kosterman and Feshbach (1989).

[24]When using binary indicators in Mplus, an IRT model is embedded within the type of measurement model estimated in this paper. I do not report the parameter estimates from the IRT model, since these occurred in only two instances.

[25]I thank Cindy Kam and Ricardo Ramírez for prodding me on this point.

assigning them to complete a battery of survey items in one language or the other. If the absence of linguistic invariance persists, one would have additional evidence tracing this pattern to the quality of translated survey items. This is the approach I am pursuing in related work on Latinos in the United States (Pérez 2009). Yet its application has relevance in linguistically diverse regions outside of the United States (e.g., Africa, Western Europe), where language may affect the nature and quality of the survey response.

Finally, though the focus of this study was the United States, the implications of this analysis are of import to survey research conducted in comparative settings, where more intricate matrices of language groups and nations magnify the main issue raised herein. Indeed, in one way, my focus on one language minority (Latinos) within one nation (United States) is a conservative estimate of a key challenge that affects survey research in linguistically diverse areas. In regions such as Africa and Latin America, scholars may uncover more complexity as they examine how language affects survey data within single nations and across nations sharing variants of a common language. The higher-order magnitude of linguistic diversity in these regions lends itself to analyses that can explore the role of contextual effects on the linguistic validity of survey data, since language minorities are often nested within nations, which are further embedded in regional blocs. In this way, scholars may, for instance, examine the extent to which political ideologies (e.g., populism) have uniform meanings across language groups within nations. The need for scholars to create a tighter interface between the political realities they purport to measure with surveys, and the linguistic parameters of their target populations, is plain to see. And it will only grow more pressing, as regions throughout the world (e.g., North America, Western Europe) experience a reconfiguration of their linguistic landscapes by familiar forces, such as immigration.

In sum, these findings underscore the importance of diagnosing and enhancing the validity of survey items administered to linguistically diverse populations, such as the United States. As the linguistic parameters of different societies continue to expand, greater attention to the quality of survey data should help us take better stock of how much we know about the attitudes and behavior of linguistic minorities, where we have fallen short, and what directions we can take to enhance our collective understanding of these communities. The intention of this paper has been to contribute to this evolving enterprise.

Acknowledgments

I am deeply grateful for the support and critical feedback offered by the following individuals: John Aldrich, Carol Atkinson, Victoria DeFrancesco-Soto, John Geer, Kerry Haynie, Marc Hetherington, Vincent Hutchings, Cindy Kam, Alisa Kessel, Jie Lu, Paula McClain, Ricardo Ramírez, Anoop Sadanandan, Mitch Seligson, Chongming Yang, and Liz Zechmeister. Any remaining errors are my own. This research was supported by a National Science Foundation Graduate Fellowship.

Manuscript submitted 1 July 2008
Manuscript accepted for publication 15 December 2008

References

Achen, Christopher H. 1975. "Mass Political Attitudes and the Survey Response." *American Political Science Review* 69 (4): 1218–31.

Almond, Gabriel A., and Sidney Verba. 1963. *The Civic Culture*. Princeton, NJ: Princeton University Press.

Alvarez, R. Michael, and Tara L. Butterfield. 2000. "The Resurgence of Nativism in California? The Case of Proposition 187 and Illegal Immigration." *Social Science Quarterly* 81 (1): 167–80.

Andrews, Frank M. 1984. "Construct Validity and Error Components of Survey Measures: A Structural Modeling Approach." *Public Opinion Quarterly* 48 (2): 409–42.

Asher, Herbert B. 1974. "Some Consequences of Measurement Error in Survey Data." *American Journal of Political Science* 18 (2): 469–85.

Benet-Martinez, Veronica, and Oliver P. John. 1998. "*Los Cinco Grandes* Across Cultures and Ethnic Groups: Multitrait Multimethod Analyses of the Big Five in Spanish and English." *Journal of Personality and Social Psychology* 75 (3): 729–50.

Bobo, Lawrence, James Johnson, Melvin Oliver, Reynolds Farley, Barry Bluestone, Irene Browne, Sheldon Danziger, Gary Green, Harry Holzer, Maria Krysan, Michael Massagli, and Camille Zubrinsky Charles. 1992–1994. *Multi-City Study of Urban Inequality: Atlanta, Boston, Detroit, and Los Angeles*. Ann Arbor, MI: Inter-University Consortium for Social and Political Research.

Bobo, Lawrence, and Devon Johnson. 2000. "Racial Attitudes in a Prismatic Metropolis: Mapping Identity, Stereotypes, Competition, and Views on Affirmative Action." In *Prismatic Metropolis: Inequality in Los Angeles*, eds. Lawrence Bobo, Melvin L. Oliver, James H. Johnson, and Abel Valenzuela. New York: Russell Sage Foundation, 81–163.

Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.

Brislin, Richard W. 1970. "Back-Translation for Cross-Cultural Research." *Journal of Cross-Cultural Psychology* 1 (3): 185–216.

Browne, Michael W., and Robert Cudeck. 1993. "Alternative Ways of Assessing Model Fit." In *Testing Structural Equation*

Models, eds. Kenneth A. Bollen and J. Scott Long. Newbury Park, CA: Sage, 136–62.

Byrne, Barbara M., Richard J. Shavelson, and Bengt Muthén. 1989. "Testing for the Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance." *Psychological Bulletin* 105: 456–66.

Chan, Bibiana, Gordon Parker, Lucy Tully, and Maurice Eisenbruch. 2007. "Cross-Cultural Validation of the DMI-10 Measure of State Depression: The Development of a Chinese Language Version." *Journal of Nervous Mental Disorders* 195: 20–25.

Citrin, Jack, Beth Reingold, and Donald P. Green. 1990. "American Identity and the Politics of Ethnic Change." *Journal of Politics* 52 (4): 1124–54.

Citrin, Jack, Donald P. Green, Christopher Muste, and Cara Wong. 1997. "Public Opinion Toward Immigration Reform: The Role of Economic Motivations." *Journal of Politics* 59 (3): 858–81.

Citrin, Jack, Cara Wong, and Brian Duff. 2001. "The Meaning of American National Identity: Patterns of Ethnic Conflict and Consensus." In *Social Identity, Intergroup Conflict, and Conflict Reduction*, eds. Richard D. Ashmore and Lee J. Jussim, and David Wilder. London: Oxford University Press, 71–100.

Citrin, Jack, Amy Lerman, Michael Murakami, and Kathryn Pearson. 2007. "Testing Huntington: Is Hispanic Immigration a Threat to American Identity?" *Perspective on Politics* 5 (1): 31–48.

Converse, Phillip E. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent*, ed. David E. Apter. New York: Free Press, 206–61.

Davis, Darren W., and Ronald E. Brown. 2002. "The Antipathy of Black Nationalism: Behavioral and Attitudinal Implications of an African American Ideology." *American Journal of Political Science* 46 (2): 239–52.

Dawson, Michael C. 1994. *Behind the Mule: Race and Class in African-American Politics*. Chicago: University of Chicago Press.

Dawson, Michael C. 2001. *Black Visions: The Roots of Contemporary African-American Political Ideologies*. Chicago: University of Chicago Press.

de Figueiredo, Rui J. P., and Zachary Elkins. 2003. "Are Patriots Bigots? An Inquiry into the Vices of In-Group Pride." *American Journal of Political Science* 47 (1): 171–88.

de la Garza, Rodolfo O., Angelo Falcon, F. Chris Garcia, and John A. Garcia. 1989–1990. *Latino National Political Survey*. Ann Arbor, MI: Inter-University Consortium for Social and Political Research.

de la Garza, Rodolfo O., Jerry L. Polinard, Robert D. Wrinkle, and Tomas Longoria. 1991. "Understanding Intra-Ethnic Attitude Variations: Mexican Origin Population Views of Immigration." *Social Science Quarterly* 72 (2): 379–87.

de la Garza, Rodolfo O., F. Chris Garcia, and Angelo Falcon. 1992. *Latino Voices: Mexican, Puerto Rican, and Cuban Perspectives on American Politics*. Boulder, CO: Westview Press.

de la, Garza, Angelo Falcon, F. Chris Garcia. 1996. "Will the Real Americans Please Stand Up: Anglo and Mexican-American Support of Core American Values." *American Journal of Political Science* 40 (2): 335–51.

DeShon, R. P. 1998. "A Cautionary Note on Measurement Error Corrections in Structural Equations." *Psychological Methods* 4: 192–211.

Diaz-Morales, Juan Francisco, Joseph R. Ferrari, Karem Diaz, and Doris Argumedo. 2006. "Factorial Structure of Three Procrastination Scales with a Spanish Adult Population." *European Journal of Psychological Assessment* 22 (2): 132–37.

Ervin, Susan, and Robert T. Bower. 1952–53. "Translation Problems in International Surveys." *Public Opinion Quarterly* 16 (4): 595–604.

Feshbach, Seymour. 1991. "Attachment Processes in Adult Political Ideology: Patriotism and Nationalism." In *Intersections with Attachment* eds. Jacob L. Gewirtz and William M. Kurtines. Hillsdale, NJ: Lawrence Erlbaum Associates, 207–26.

Fetzer, Joel S. 2000. *Public Attitudes Toward Immigration in the United States, France, and Germany*. Cambridge, MA: Cambridge University Press.

Finney, Sara J., and Christine DiStefano. 2006. "Non-normal and Categorical Data in Structural Equation Modeling." In *Structural Equation Modeling: A Second Course*, eds. Gregory R. Hancock and Ralph O. Mueller. Greenwich, CT: Information Age Publishing, 269–314.

Garcia, John A. 2003. *Latino Politics in America: Community, Culture, and Interests*. Lanham, MD: Rowman & Littlefield Publishers, Inc.

Gilens, Martin. 1996. "'Race Coding' and White Opposition to Welfare. *American Political Science Review* 90 (3): 593–604.

Gilens, Martin. 1999. *Why Americans Hate Welfare: Race, Media, and the Politics of Antipoverty Policy*. Chicago: University of Chicago Press.

Hellwig, David J. 1982. "Strangers in Their Own Land: Patterns of Black Nativism." *American Studies* 23: 85–98.

Higham, John. 1955. *Strangers in the Land: Patterns of American Nativism, 1860–1925*. New Brunswick, NJ: Rutgers University Press.

Hood, M. V., Irwin L. Morris, and Kurt A. Shirkey. 1997. "Quedate o Vente: Uncovering the Determinants of Hispanic Public Opinion toward Immigration." *Political Research Quarterly* 50 (3): 627–47.

Hood, III, M. V., and Irwin L. Morris. 1997. "Amigo o Enemigo? Context, Attitudes, and Anglo Public Opinion toward Immigration." *Social Science Quarterly* 78 (2): 309–23.

Hoyle, Rick H. 1995. "The Structural Equation Modeling Approach: Basic Concepts and Fundamental Issues." In *Structural Equation Modeling: Concepts, Issues, and Applications*, ed. R. H. Hoyle. Thousand Oaks, CA: Sage Publications, Inc.

Huddy, Leonie, and Nadia Khatib. 2007. "American Patriotism, National Identity, and Political Involvement." *American Journal of Political Science* 51 (1): 63–77.

Hurwitz, Jon, and Mark Peffley. 1987. "How are Foreign Policy Attitudes Structured? A Hierarchical Model." *American Political Science Review* 81 (4): 1099–1120.

Ignatiev, Noel. 1995. *How The Irish Became White*. New York: Routledge.

Jacobson, Mathew Frye. 1998. *Whiteness of a Different Color: European Immigrants and the Alchemy of Race*. Cambridge, MA: Harvard University Press.

Jacobson, Eugene, Hideya Kumata, and Jeanne E. Gullahorn. 1960. "Cross-Cultural Contributions to Attitude Research." *Public Opinion Quarterly* 24 (2): 205–23.

Kaplan, David. 2000. *Structural Equation Modeling: Foundations and Extensions*. Thousand Oaks, CA: Sage Publications, Inc.

Kessler, Alan. 2001. "Immigration, Economic Insecurity, and the "Ambivalent" American Public." *Center for Comparative Immigration Studies Working Paper*. No. 41: 1–50.

Kim, Jae-On, and Charles W. Mueller. 1978. *Introduction to Factor Analysis: What It Is and How to Do It*. Thousand Oaks, CA: Sage Publications, Inc.

King, Desmond. 2000. *Making Americans: Immigration, Race, and the Origins of the Diverse Democracy*. Cambridge, MA: Harvard University Press.

King, Gary, Christopher J. L. Murray, Joshua A. Salomon, and Ajay Tandon. 2004. "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research." *American Political Science Review* 98 (1): 191–207.

Kline, Rex B. 2005. *Principles and Practice of Structural Equation Modeling*. New York: The Guilford Press.

Kosterman, Rick, and Seymour Feshbach. 1989. "Toward a Measure of Patriotic and Nationalistic Attitudes." *Political Psychology* 10 (2): 257–74.

Lambert, Michael Canute, Neal Schmitt, Maureen E. Samms-Vaughan, Jeong Shin An, Maureen Fairclough, and Christine A. Nutter. 2003. "Is It Prudent to Administer All Items for Each Child Behavior Checklist Cross-Informant Syndrome? Evaluating the Psychometric Properties of the Youth Self-Report Dimensions with Confirmatory Factor Analysis and Item Response Theory." *Psychological Assessment* 15 (4): 550–68.

Lee, Taeku. 2002. "Language-of-Interview Effects and Latino Mass Opinion." Presented at the annual meeting of the Midwest Political Science Association.

MacCallum, Robert C. 1995. "Model Specification: Procedures, Strategies, and Related Issues." In *Structural Equation Modeling: Concepts, Issues, and Applications*, ed. R. H. Hoyle. Thousand Oaks, CA: Sage Publications, Inc., 16–36.

Masuoka, Natalie. 2006. "Together They Become One: Examining the Predictors of Panethnic Group Consciousness among Asian Americans and Latinos." *Social Science Quarterly* 87 (5): 993–1011.

McClain, Paula D., Niambi M. Carter, Victoria M. DeFrancesco Soto, Monique L. Lyle, Jeffrey D. Grynaviski, Shayla C. Nunally, Thomas J. Scotto, J. Alan Kendrick, Gerald F. Lackey, and Kendra Davenport-Cotton. 2006. "Racial Distancing in a Southern City: Latino Immigrants' Views of Black Americans." *Journal of Politics* 68 (3): 571–84.

McDermott, Rose. 2002. "Experimental Methods in Political Science." *Annual Review of Political Science* 5: 31–61.

Miller, Lawrence W., Jerry Polinard, and Robert D. Wrinkle. 1984. "Attitudes toward Undocumented Workers: The Mexican American Perspective." *Social Science Quarterly* 65: 482–94.

Nicholson, Stephen, Adrian Pantoja, and Gary M. Segura. 2006. "Political Knowledge and Issue Voting Among the Latino Electorate." *Political Research Quarterly* 59 (2): 259–71.

Nobles, Melissa. 2002. *Shades of Citizenship: Race and the Census in Modern Politics*. Palo Alto, CA: Stanford University Press.

Paunonen, Sampo V., and Michael C. Ashton. 1998. "The Structured Assessment of Personality Across Cultures." *Journal of Cross-Cultural Psychology* 29 (1): 150–70.

Pérez, Efrén O. 2009. "Speaking in Tongues: Language, Politics, and the Survey Response." Typescript. Vanderbilt University.

Perreira, Krista M., Natalia Deeb-Sossa, Kathleen Mullan Harris, and Kenneth Bollen. 2005. "What Are We Measuring? An Evaluation of the CES-D across Race/Ethnicity and Immigrant Generation." *Social Forces* 83 (4): 1567–1602.

Reise, Steven P., Keith F. Widaman, and Robin H. Pugh. 1993. "Confirmatory Factor Analysis and Item Response Theory: Two Approaches for Exploring Measurement Invariance." *Psychological Bulletin* 114: 552–66.

Sanchez, Gabriel R. 2006. "The Role of Group Consciousness in Latino Public Opinion." *Political Research Quarterly* 59 (3): 435–46.

Saris, Willem E., and Frank M. Andrews. 1991. "Evaluation of Measurement Instruments Using a Structural Equations Modeling Approach." In *Measurement Errors in Surveys*, ed. Paul B. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman. New York: Wiley, 575–98.

Sears, David O., Colette Van Laar, Mary Carrillo, and Rick Kosterman. 1997. "Is It Really Racism? The Origins of White Americans' Opposition to Race-Targeted Policies." *Public Opinion Quarterly* 61 (1): 16–53.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Non-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.

Shankman, Arnold. 1982. *Ambivalent Friends: Afro-Americans View the Immigrant*. Westport, CT: Greenwood Press.

Shingles, Richard. 1981. "Black Consciousness and Political Participation: The Missing Link." *American Political Science Review* 75: 76–91.

Spini, Dario. 2003. "Measurement Equivalence of 10 Value Types From the Schwartz Value Survey Across 21 Countries." *Journal of Cross-Cultural Psychology* 34 (1): 3–23.

Stern, Eric. 1948–49. "The Universe, Translation, and Timing." *Public Opinion Quarterly* 12 (4): 711–15.

Stimson, James A. 1999. *Public Opinion in America: Moods, Cycles, and Swings*. Boulder, CO: Westview Press.

Tate, Katherine. 1994. *From Protest to Politics: The New Black Voters in American Elections*. Cambridge, MA: Harvard University Press and the Russell Sage Foundation.

Tolbert, Caroline J., and Rodney E. Hero. 1996. "Race/Ethnicity and Direct Democracy: An Analysis of California's Illegal Immigration Initiative." *Journal of Politics* 58 (3): 806–18.

Treier, Shawn, and Simon Jackman. 2008. "Democracy as a Latent Variable." *American Journal of Political Science* 52 (1): 201.

Tourangeau, Robert, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge, MA: Cambridge University Press.

Uhlaner, Carole J., and F. Chris Garcia. 2002. "Latino Public Opinion." In *Understanding Public Opinion*, eds. Barbara Norrander and Clyde Wilcox. Washington, DC: CQ Press.

U.S. Census Bureau. 2007. *Statistical Abstract of the United States*. Washington, DC.

van der Veld, William, and Willem E. Saris. 2004. "Separation of Error, Method Effects, Instability, and Attitude Strength." In *Studies in Public Opinion: Gauging Attitudes, Non-Attitudes, Measurement Error, and Change*, eds. Willem E. Saris and Paul M. Sniderman. Princeton, NJ: Princeton University Press, 37–59.

Welch, Susan, John Comer, and Michael Steinman. 1973. "Interviewing in a Mexican-American Community: An Investigation of Some Potential Sources of Response Bias." *Public Opinion Quarterly* 37 (1): 115–26.

Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. Cambridge, MA: Cambridge University Press.

Efrén O. Pérez is assistant professor of political science, Vanderbilt University, Nashville, TN 37235.