The Position-Reputation-Information (PRI) scale of individual prestige

Richard E.W. Berl[1*], Alarna N. Samarasinghe[2], Fiona M. Jordan[2,3], Michael C. Gavin[1,3]

[1] Department of Human Dimensions of Natural Resources, Colorado State University, Fort Collins, Colorado, United States of America

[2] Department of Anthropology and Archaeology, University of Bristol, Bristol, United Kingdom

[3] Max Planck Institute for the Science of Human History, Jena, Germany

* Corresponding author

E-mail: rewberl@colostate.edu (REWB)

Abstract

Prestige is a key concept across the social and behavioral sciences and has been implicated as an important driver in the processes governing human learning and behavior and the evolution of culture. However, existing scales of prestige fail to account for the full breadth of its potential determinants or focus only on collective social institutions rather than the individual-level perceptions that underpin everyday social interactions. Here, we use open, extensible methods to unite diverse theoretical ideas into a common measurement tool for individual prestige. Participants evaluated the perceived prestige of regional variations in accented speech using a pool of candidate scale items generated from free-listing tasks and a review of published scales. Through exploratory and confirmatory factor analyses, we find that our resulting 7-item scale, composed of dimensions we term position, reputation, and information, or "PRI," exhibits good model fit, scale validity, and scale reliability. The PRI scale of individual prestige contributes to the integration of existing lines of theory on the concept of prestige, and the scale's application in Western contexts and its extensibility to other cultures, serves as a foundation for new theoretical and experimental trajectories across the social and behavioral sciences.

Introduction

Prestige is a key concept for many disciplines in the social and behavioral sciences, including psychology [1], sociology [2], anthropology [3], and economics [4]. Through its influence on the cultural transmission of knowledge and the dynamics that shape cultural diversity, prestige has been implicated as a crucial component in the evolution of our highly social species [5–8]. These cultural evolutionary dynamics ultimately arise from social interactions between individuals at the microevolutionary level. Therefore, we can consider the individual as the unit that acquires, holds, and benefits from prestige in day-to-day life. Despite the theoretical and practical importance of the prestige concept, it is surprising that no satisfactory tool currently exists for measuring individual prestige.

A scale of individual prestige that is theoretically and practically meaningful must have validity (e.g. it measures what it is intended to measure) and reliability (e.g. it is consistent in those measurements). When quantifying prestige, the scale must measure perceptions of the traits that constitute prestige and the relative influence these traits have on the general prestige construct. The scale should also assist researchers in accounting for differences in perceptions between groups of respondents—by culture, demographics, or otherwise—in order to avoid being misled by results from inappropriately aggregating across these groups [9–11]. In addition, the scale should be developed using replicable methods to allow for adaptations for use with new groups that may hold different values. Lastly, in developing the scale, researchers should endeavor to be data-driven and theory-neutral [12,13] to minimize the potential bias posed by researchers' expectations and to maximize the real-world utility and validity of the scale.

Rather than individual prestige, existing prestige scales focus on the prestige of collective social institutions or constructs, such as organizational prestige (regard for an institution, e.g. [14,15]), brand prestige (status associated with products, e.g. [16,17]), and occupational prestige (standing of professions, e.g. [18–20]), that are not directly derived from or attributable to individual-level traits. Some of the most widely-used "scales" of occupational prestige—including the NORC Duncan Socioeconomic Index [18], the Nakao-Treas Prestige Score [19], and the International Socio-Economic Index of Occupational Status [20] (and its predecessors, e.g. [21])—are not measurement tools, but rather lists of prior composite ratings for each occupation. Researchers obtained some of these existing prestige "scales" (and others, e.g. [22,23]) by directly asking participants to rank others by their own internal concept of prestige, left undefined, or by how participants think society in general would or should rank them. These ambiguities in previous indices of prestige leave findings open to theoretically-biased interpretations [24,25].

The distinction between data-driven and theory-driven research is also relevant when considering the suitability of another published scale for measuring individual prestige: the prestige-dominance scale developed by Cheng et al. [26]. This scale was built to conform to a specific theoretical framework [27] and contrasts "prestige" and "dominance" as opposing unidimensional constructs. To maintain theoretical soundness, Cheng and colleagues chose to retain multiple scale items that did not meet their stated inclusion criteria and contributed to a poorly-fitting final model (CFI < 0.95, GFI < 0.90, RMSEA > 0.05) [26]. Here, for the purpose of developing an accurate measurement tool, we consider that the characteristics of an individual that may contribute to prestige could also overlap with those that contribute to dominance, rather than belonging to either of two fully discrete avenues to status. Previous research [28–31] suggests that peoples' mental models for one or both of these constructs may also be multidimensional rather than unidimensional. Importantly, these hypotheses can be assessed using an empirical, theory-neutral approach.

The purpose of our work is to construct a valid and reliable scale of individual prestige, as defined by participants within two broadly "Western" societies—the United States and the United Kingdom—using replicable methods that we intend to be extensible to other contexts and cultures. We take a minimal theoretical approach, elements of which have been suggested in disparate parts of the literature but never explored together in one measurement tool. Our approach makes only three fundamental assumptions about prestige:

1. Prestige can be seen as a trait possessed and used by an individual in the course of everyday social life, distinct from but not independent of the prestige accorded to the societal institutions and constructs of which they may be a part [2,25,32];
2. Prestige is based upon the subjective assessments of others, through the lens of their individually, socially, and culturally acquired beliefs, values, attitudes, and experiences [2,3,25,28,33,34]; and
3. Prestige may be composed of multiple dimensions [2,28–31,35,36], each representing differential contributions from individual, social, or cultural domains.

We made no further assumptions about what constitutes prestige or of its specific societal mechanisms and consequences, as our goal was to obtain the necessary information from respondents' own views of prestige in the real world [25]. Our approach was driven to a large degree by the responses of participants, rather than relying on any specific, theoretically-entrenched prestige concept.

One methodological challenge of our approach involved finding a valid, widely-recognized signal of prestige that could be presented to participants to evaluate the pool of prospective prestige scale items. Ideally, this instrument would also avoid pre-defining for participants what prestige means. For this purpose, and because this is one component of a larger study on prestige and the transmission of spoken information, we chose to use accented regional variation in speech to highlight differences in individual prestige. Work by sociolinguists has consistently shown that linguistic characteristics such as dialect and accent can index macro-social categories related to prestige (such as class) in the perceptions of listeners, as well as acquiring socially significant meanings of their own. Accents and regional varieties are therefore perceived as strong indicators of prestige and tend to be stable over time [37–40]. Accents are hard-to-fake signals [41] and because accents that are regarded as locally "standard" or associated with desirable upper class membership tend to be evaluated highly by a majority of listeners, they often serve as an index of membership in a high-status group [37,42,43]. Naturally, some disagreement will exist between different demographic groups on the evaluation of particular accents [37,44]. However, our focus is not on how respondents rate specific accents but on the relationships between the items used in the evaluation of prestige.

The development of a valid and reliable scale will enable researchers from diverse disciplinary backgrounds to measure individual prestige using a shared prestige concept. The scale can thus contribute to the evaluation and reconciliation of competing theories on prestige and serve as a foundation for the development of new theoretical and experimental trajectories across the social and behavioral sciences.

Results

The scale development process involved first constructing the prospective scale by collecting items and determining their structure through exploratory factor analysis, then evaluating the fit of the model using confirmatory factor analysis with a separate data set, and finally assessing the validity and reliability of the scale using a mixture of qualitative and quantitative criteria.

Study 1: Scale construction

We began by conducting a study to generate a pool of words or phrases ("items") related to prestige, reducing the items to those most indicative of prestige, and constructing the scale by establishing the factor structure of those items using exploratory factor analysis ("EFA"). We collected items from three sources: the most salient terms in a free-listing task completed by participants; a previously unpublished pilot study on sociolinguistic prestige; and a review of published scales that measure language attitudes and incorporated a prestige or status dimension. We also collected items from two contrasting domains—“solidarity” and “dynamism”—from published sources, to ensure that scale items adequately discriminated between prestige and other unrelated concepts with positive connotations. We used the resulting list of items (**Table 1**) for this study and for the follow-up scale evaluation study.

Table 1. Pool of attitudinal items retained and used in the scale construction and scale evaluation studies. Reversed items used in the scale evaluation study are noted parenthetically.

PRESTIGE	SOLIDARITY	DYNAMISM
prestigious	*friendly*	*aggressive*
wealthy	*kind (unkind)*	*active*
high social status	*good-natured*	*confident*
powerful	*warm*	*enthusiastic*
respected	*comforting*	
educated		
hardworking		
successful		
intelligent (unintelligent)		
reputable		
ambitious (unambitious)		

We recruited participants from the US (n = 153) and UK (n = 155) to complete an online survey using these items to evaluate the characteristics of four speakers with varying regional accents of English. As a second complementary source of data on perceptions of association between items without involving accents, participants were also asked to group the prestige domain items into like and unlike categories using a triad test.

By sequentially applying EFA and eliminating items that failed to reach the predetermined acceptance criteria (see Methods), we obtained the best-supported factor structure for the attitudinal items across all three domains (**Table 2A** and **Fig 4**), as well as the internal factor structure of the attitudinal and triad items in the prestige domain (**Fig 1**; **Table 2B** and **2C**). Using EFA, items within the prestige domain were partitioned into three factors: *wealthy*, *powerful*, and *high social status* in the first factor, hereafter referred to as "position"; *reputable* and *respected* in the second factor, referred to as "reputation"; and *educated* and *intelligent* in the third factor, referred to as "information." We therefore denote the resulting factor structure as Position-Reputation-Information, or "PRI." Subsequent cluster analyses on the same data generated clusters that matched the three PRI factors (**Fig 6A**), as did results from comparable analyses of the triad data (**Fig 6B**), supporting the robustness of this structure.

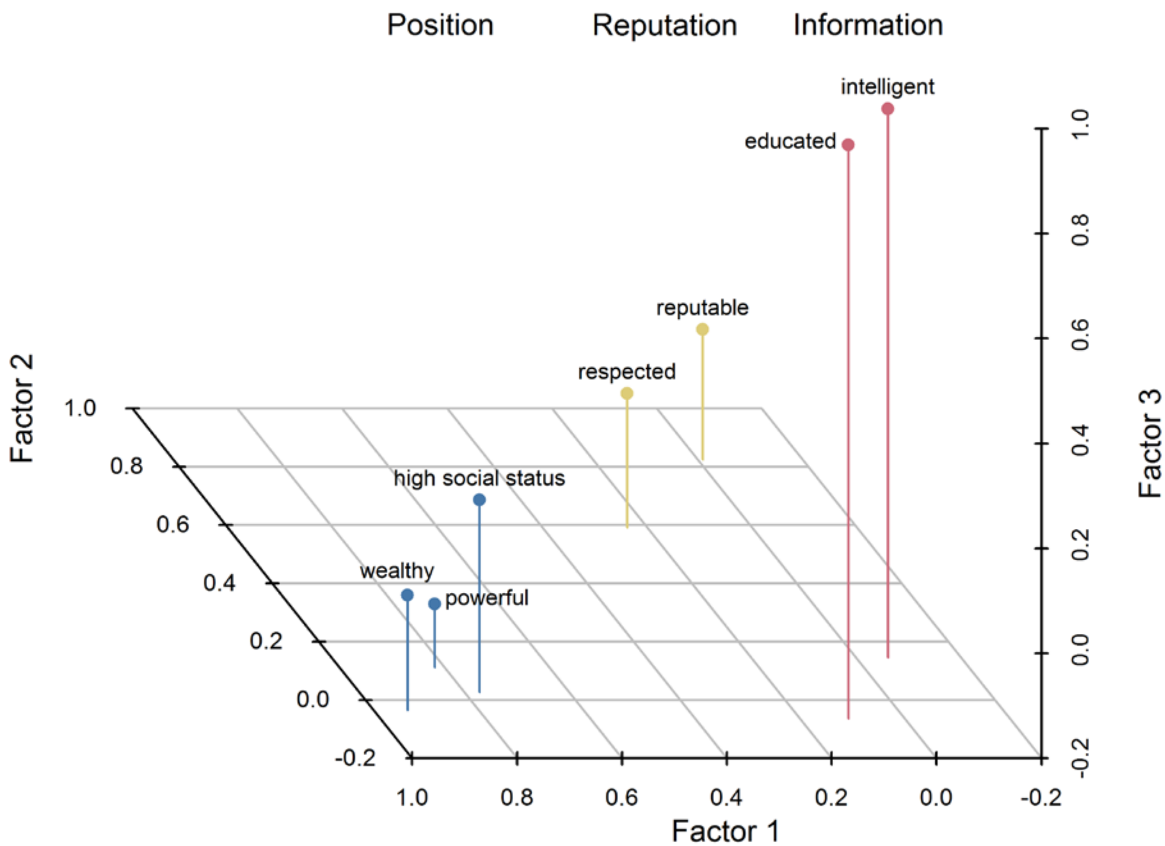


Fig 1. Prestige domain item loadings from exploratory factor analysis of attitudinal data. Position, reputation, and information items are shown in light blue, gold, and pink, respectively.

Study 2: Scale evaluation

We then conducted a second study with an independent data set to validate the findings of the scale construction study using confirmatory factor analysis (“CFA”). The validation step evaluates the fit of the structural model proposed by EFA and examines any potential systematic variance due to sampling [45]. We used the full set of relevant items from the scale construction study in the CFA, with three items presented in reversed form to reduce potential response bias (but this was found to be ineffective, see Methods).

For this study, we recruited a new, independent sample of participants from the US (n = 151) and UK (n = 144) to provide attitudinal ratings for a greater variety of accented speakers than in the previous study (n = 8 in each country, 4 of which were presented to participants in both countries; see **Table 3**), again using an online survey.

After controlling for potential differences between participant demographics, we found that the PRI model exhibited good fit (CFI = 0.959, TLI = 0.983, RMSEA = 0.031 [90% CI: 0.026, 0.036], SRMR = 0.023). Following this validation by CFA, we obtained the complete PRI scale (**Fig 2**).

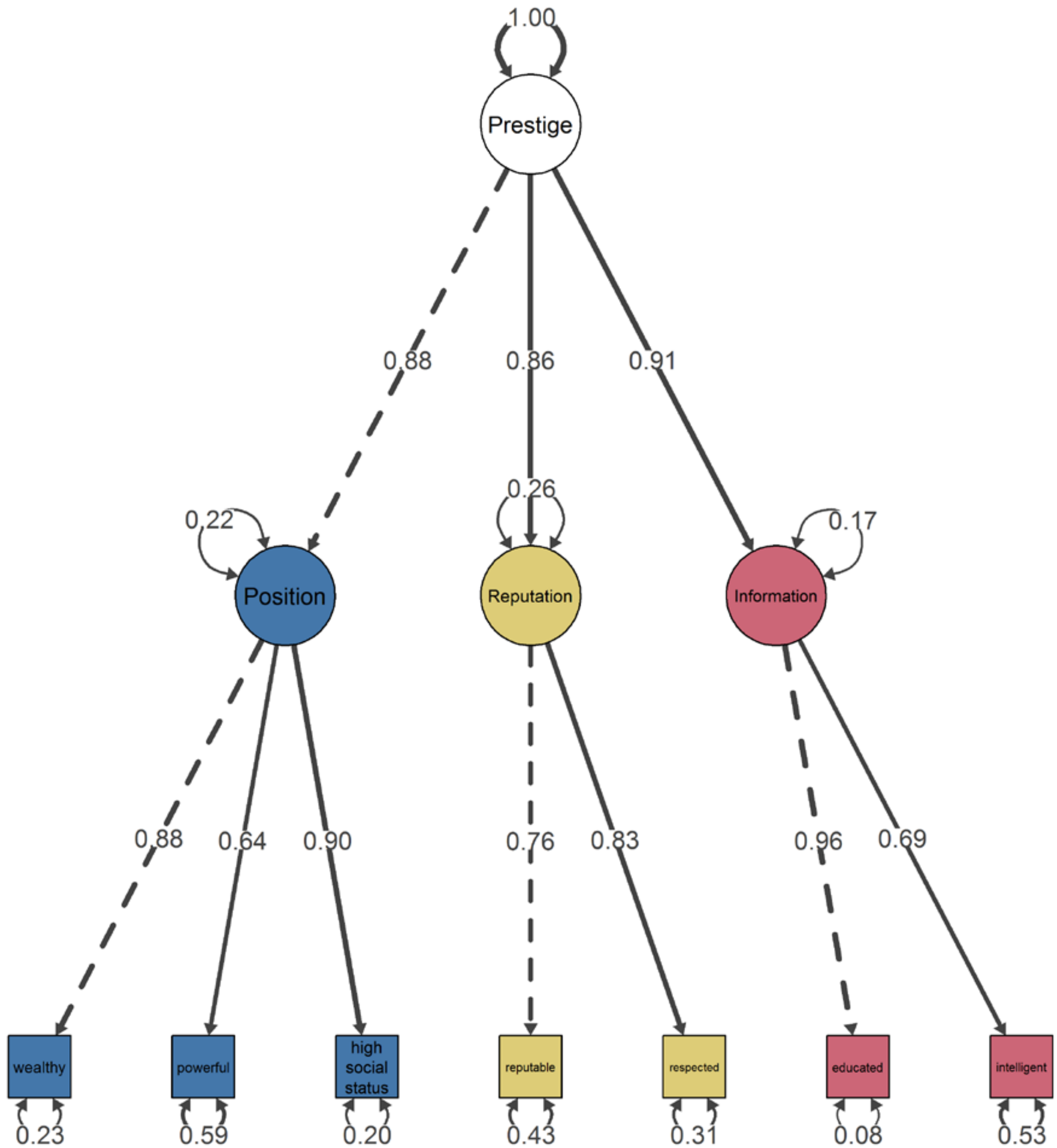


Fig 2. Path diagram and estimates from confirmatory factor analysis of the Position-Reputation-Information scale model. Standardized parameter estimates are shown as weighted edges. Residual variances are shown as self-loops. Dotted lines indicate that the loadings of the first indicator of each factor were fixed to 1.0 for estimation.

Scale validity and reliability

The PRI scale displayed both validity and reliability in the context of our samples. Using predetermined criteria to judge the acceptability of each index (see Methods), we found support for the components of construct validity: convergent validity measures exceeded the criterion for all subscales (average variance explained, or "AVE": position = 0.670, reputation = 0.629, information = 0.696) and discriminant validity measures (heterotrait-monotrait ratio, or "HTMT": **Table 6**) remained below the threshold in all cases except in one comparison between internal position and information subscales. Reliability measures of internal consistency (coefficients alpha and omega: **Table 7**) were high within each PRI subscale (M = 0.813, SD = 0.036) and for the scale as a whole (M = 0.892, SD = 0.018). Criterion validity was demonstrated by high correlations between scale items and a separate *prestigious* item (M = 0.692, SD = 0.097). As added support for the criterion validity of the PRI scale, in a comparative data set the factor scores predicted by the PRI scale were highly correlated with those of the prestige factor of the Cheng et al. [26] prestige-dominance scale (PRI overall: 0.850, position: 0.805, reputation: 0.861, information: 0.828) and the PRI scale displayed better model fit overall (ΔCFI = 0.025, ΔTLI = 0.029, ΔRMSEA = -0.045, ΔSRMR = -0.064; see Methods).

These assessments demonstrate that the PRI scale adequately represents the prestige construct and that it is distinct from the other positive traits tested (i.e. solidarity and dynamism). The three subscales (position, reputation, and information) represent cohesive parts of a whole while being relatively distinct from one another. Additionally, perceptions of the PRI structure were consistent across respondents and the scale compares well with existing prestige concepts. We take these results as support for the PRI scale as the most accurate and realistic reflection of our participants' internal views on the content and structure of the individual prestige construct.

Discussion

In the process of developing the PRI scale, we intentionally minimized the role of theory and allowed the structure inherent in the data—structure provided by participants' own internal conceptions of prestige and revealed through exploratory factor analysis—to dictate what was most relevant. However, in examining this structure and the constituent items of the scale after its formation, we found that the PRI prestige construct is highly consistent with different streams of prior research on prestige. The terms chosen to represent the three subscales, "position," "reputation," and "information," characterize three relatively distinct axes of individual prestige, and we examine each in turn.

The position components of the scale signify an individual's relative place in the social hierarchy, determined to a large extent by the circumstances of birth, family, and inheritance. Max Weber, in his classic theory of social stratification, argued that one's social position can be attributed to three dimensions: economic "class," or wealth; "status," or honor gained through prestige; and "party," or political power and influence [46,47]. These closely mirror the three items found in the position subscale (*wealthy*, *high social status*, and *powerful*) and this finding reflects the continuing utility of Weber's ideas in sociological theory and practice [48].

The items in the reputation subscale (reputable and respected) relate to social opinion and esteem and are terms frequently used to describe prestige (e.g. [14,15,49]), and are even used synonymously with it (e.g. [50]). In the sociological literature on prestige, reputation and respect have the connotation of a collective judgment of character independent of individual variation in judgments [2]. Reputation and respect represent the general societal evaluation of an individual in a certain position or role, subjectively interpreted through social and cultural values. By contrast, the items in the position subscale may be established through privilege without necessarily undergoing the same degree of collective evaluation [46,47].

The third subscale, information, and its items (*educated* and *intelligent*) represent the value placed by society on the holders of wisdom, expertise, and learning. These constructs are supported by the occupational prestige literature, which emphasizes that—in a stratified society with specialized occupations—an individual's educational background and achievement are highly predictive of their future occupational class which, in turn, contributes significantly to individual prestige (e.g. [51–53]). The salience of this subscale and its focus on information holders could also indicate support for arguments from information theory about the evolution of prestige and its role in cultural transmission. The information theory-based account, presented alongside (but not integral to) the dichotomous prestige-dominance distinction by Henrich & Gil-White [27], asserts that individuals gain prestige by having desirable skills and knowledge that others compete within a social group for the opportunity to learn. Alternatively, an occupation attained through greater education could be another avenue to wealth and power. This question, and to what extent—if any—some form of the information subscale would be relevant to prestige across the diversity of non-Western or non-industrialized societies remains open to future study.

Indeed, there is a great need to explore concepts of prestige cross-culturally to reach beyond the perspectives given by Western and westernized participants. Many existing prestige indices have been explicitly promoted for their universality, in spite of having been developed using data almost exclusively from “WEIRD” (Western, Educated, Industrialized, Rich, and Democratic) societies [54] in the 1960s, ‘70s, and ‘80s. The utility of these indices across cultures and over the significant span of time and sociocultural change that has occurred since they were developed has been called into question [9–11,28,55].

The concept of prestige, the individual components that comprise prestige, the degree of importance attached to each component, and the relationships between components are all—to some degree—culturally constructed and malleable through cultural evolutionary processes. Therefore, we recognize that the PRI scale is not universally applicable, as this is an unrealistic expectation. We developed the PRI scale using data collected from adults in the highly WEIRD societies of the United States and United Kingdom and it should not be generalized beyond that context without adequate validation. The high degree of consistency in the PRI structure across our representative samples of demographically diverse participants in the US and UK suggests that the PRI scale should function well across other highly Westernized, English-speaking societies. However, distinct demographic or cultural groups within these societies may hold different values and have substantially different internal models of prestige. For these reasons, and in the interest of following best practices in psychometrics [56], we strongly recommend testing the validity and reliability of the

PRI scale with each application and testing for invariance across as many demographic variables as may be potentially relevant.

We have made the process of constructing and validating the PRI scale extensible to any additional population for which a scale of individual prestige is needed, through the emphasis on the participants in the item generation and evaluation stages, the use of straightforward and appropriate methods and criteria, the use of open-source analytical tools, and the open sharing of all data and code used to run analyses (see Supplementary Materials). A new variant of the PRI scale can be constructed by repeating these methods in a new group, with awareness and care for local cultural norms and power structures. Examining systematic differences in responses and extending the PRI scale to other contexts and cultures can further improve the representation and inclusion of minority and non-Western perspectives on prestige, and we argue is the most important avenue for future research presented by this study.

The PRI scale for the measurement of individual prestige fills a crucial niche by establishing a measurement tool driven by the real-world perceptions of individuals across two Western societies. The PRI scale enables the study of prestige—a central yet divisive concept throughout the social and behavioral sciences—using a common foundation, which we hope will encourage fruitful engagement, conversation, and collaboration that spans across disciplinary boundaries. We have shown the broad utility of this scale for conducting research by finding support for the PRI structure in both of two separate sources of data: attitudinal responses to variations in accented speech, and triadic conceptual associations absent the sociolinguistic context.

Future research should endeavor to untangle the complex and varied patterns in how prestige is perceived and how it operates in the practice of real social interactions across the breadth of human experience. The availability of the PRI scale allows researchers to explore in greater detail the relationships between different aspects of prestige, dominance, status, and success. Some of these relationships may be quite complex, or even circular, as suggested by the presence of *high social status* as an indicator of prestige within the position subscale (whereas scholars would normally consider prestige to be a contributor to status) or by the possible contributions of specific indicators like *educated* toward other indicators like *wealthy*. Additionally, there may be some degree of overlap between the construct of prestige, as measured by the PRI scale and the prestige factor of the Cheng et al. [26] prestige-dominance scale, and other related concepts like dominance and leadership. Many questions remain about the breadth and interconnectedness of the varied routes to the acquisition of social status. We view the establishment of the PRI scale as a necessary step toward a more integrated and comprehensive understanding of prestige, through the clarification of preceding debates and the beginning of new lines of inquiry into the core concepts that shape interactions, relationships, social structure and inequality, and the evolution of culture.

Methods

Ethics statement

We obtained written prior informed consent from all participants in this research. Participants that completed surveys through the Amazon Mechanical Turk and Prolific platforms were compensated above hourly minimum wage, in the state of Colorado for US participants and in the UK for participants located there, based upon the time needed to complete the surveys. Participants self-reported demographic information for socioculturally determined constructs such as ethnicity and gender, using categories in accordance with current local and ethical guidelines. Full details on these categories are given in the description of each data set in the Supplementary Materials.

Prior approval for research protocols was obtained from the Colorado State University Institutional Review Board (protocol #014-16H) and the University of Bristol

Faculty of Arts Human Research Ethics Committee (protocols #26561, #31041, and #38323).

Study 1: Scale construction

Item generation

In the development of this scale, we used a combination of deductive and inductive methods to collect the items most relevant to the concept of individual prestige. This methodological approach incorporated emic, operational determinants of prestige from a real-world Western context, as well as shared items from previous scales, in order to evaluate all possible components of a prestige scale concurrently. We sampled items from a salience analysis of responses to a free listing task, from existing attitudinal scales in the literature, and from responses to a pilot study investigating sociolinguistic prestige. We favored the use of inductive methods, specifically the free listing task, because they are generalizable and facilitate replication and extension to other contexts and cultures.

Free listing is a tool from cultural domain analysis used to elicit responses on a particular classification of knowledge [57–59]. The task conducted as part of this study consisted of a survey in which participants responded to the following three prompts, in order:

1. *List all of the words or phrases that you can think of that are related to "prestige."*
2. *List all of the words or phrases that you can think of that describe "prestigious" people.*
3. *List all of the characteristics that you can think of that make a person "prestigious."*

Responses were limited to 2 minutes per question. We allowed repetition of terms from prior questions, but participants could not refer back to previous responses. We recruited participants for this task through advertisements in local undergraduate courses (n = 6 US) and social media networks (n = 42 US, 20 UK), for a final sample of 68 participants. We compensated undergraduate students for their participation and social media participants engaged voluntarily. Participants ranged in age from 18 to 50 (M = 28.9, SD = 7.3), with 18 that identified as male and 50 that identified as female. All participants were native English speakers. All participants self-identified as white, except for one person of mixed ethnicity from the US and one person of color from the UK. Participants came from a variety of backgrounds with respect to the size of their childhood settlement and educational attainment, but for occupation most were either students (25.0%) or were in management or professional positions (26.5%), with others in service (11.8%) and sales (5.9%) positions.

After obtaining the three free lists of items from each participant, we grouped items for common meaning, reducing the pool of unique items from 717 to 303. Generally, this procedure consisted of replacing multi-word phrases with single-word synonyms and converting words to adjective form (e.g. "lots of education" to "educated" and "influence" to "influential"). We left given terms as-is if their intended meaning was ambiguous. On the whole, groupings were done with the intent of minimal replacement, so as to allow participants to speak for themselves, and all co-authors verified the groupings. We then calculated a salience value for each of the 303 items using Smith's S [60], which takes into account both the frequency of an item's occurrence across lists and order of occurrence within lists. From a scree plot of the items by their salience values, we chose the cutoff near the inflection point at the highest local proportional drop in salience (0.0148) to capture the set of most salient items (**Fig 3**) [57]. The items retained from this exercise were: *wealthy*, *high social status*, *powerful*, *respected*, *educated*, *hardworking*, and *successful*.

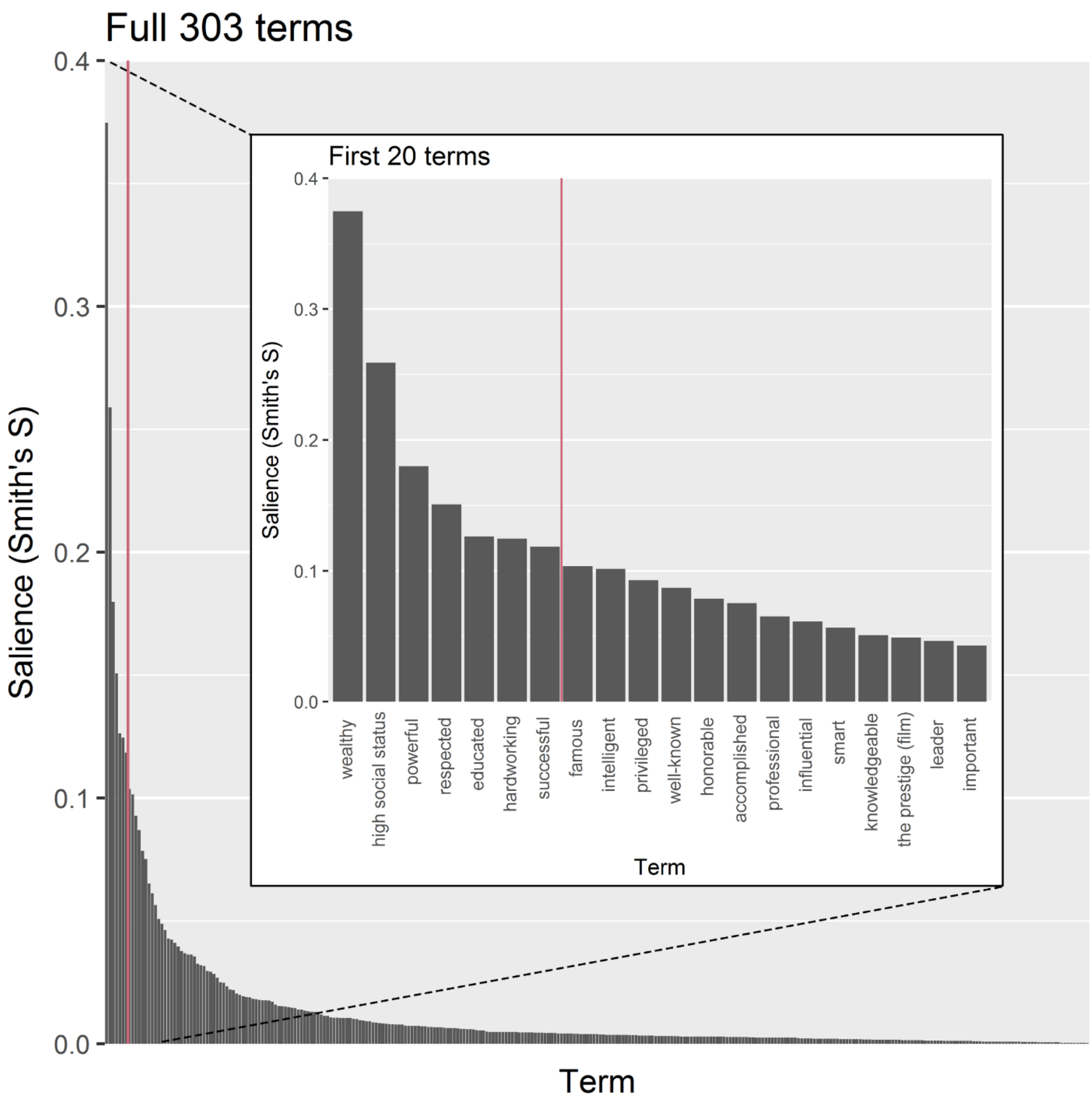


Fig 3. Scree plot of free list items by salience value. Chosen cutoff includes *successful* (S = 0.1185) and items of higher salience and excludes *famous* (S = 0.1037) and items of lower salience.

Given the use of attitudes toward regional accents as a measurement tool in this study, and in the interest of full coverage of the domain of interest (i.e. content validity), we chose to supplement the pool of potential scale items by reviewing items used in established scales of language attitudes that incorporated a prestige or status dimension. The two scales we selected for this purpose were the Speech Dialect Attitudinal Scale ("SDAS": [61]) and its revised version ("SDAS-R": [62]) and the Speech Evaluation Instrument ("SEI": [63]) and its short form version ("SEI-S," as used by [64]). The following items were represented in some form within both scales under a dimension of "prestige," "status," or "competence" and were therefore retained: *wealthy*, *high social status*, *educated*, and *intelligent* (as a note, we collapsed *upper class* into the broader *high social status* and *literate* into *educated*). These items agreed closely with those used in other sociolinguistic studies for these dimensions [40], and therefore can be regarded as representative of the literature. We also collected items from the selected scales to represent two other domains commonly used in speech evaluation studies [37,40]: "solidarity" and "dynamism." We included these domains, which are unrelated to prestige but similarly positively valenced, to assess the ability of prestige items to represent prestige itself and not merely a positive evaluation of the speaker (i.e. discriminant validity). The additional items we selected were: *friendly*, *kind*, *good natured*, *warm*, and *comforting* for the "solidarity" dimension, and *aggressive*, *active*, *confident*, and *enthusiastic* for the "dynamism" dimension. One item, *clear*, was also initially included within "dynamism," but we later removed it from analyses due it clustering more closely with items in other dimensions.

A third and final source of items was a previously unpublished pilot study that we conducted on speech, accent, and prestige in October and November of 2015. The sample of this pilot study consisted of 100 US and 44 UK participants (undergraduate and graduate/postgraduate students) ranging in age from 18 to 64 years (M = 21.8, SD = 5.3). Of the participants, 47 identified as male and 95 as female. The majority of participants, 141, identified as native English speakers, with 2 non-native speakers. Participants were asked to rate two speakers—one with a locally standard accent (US or UK) and the other with a nonstandard accent—on 15 attitudinal items using a 7-point Likert-type scale. The items in this pilot study were also drawn from prior linguistic studies. Following exploratory factor analysis and the sequential elimination of items following the same criteria described below for the present study, as well as examining inter-item correlations with a *prestigious* item to find the most closely associated items, we retained the following items from the pilot study: *hardworking*, *reputable*, *intelligent*, and *ambitious*.

The combined pool of items retained from all three sources (**Table 1**) were then used in the scale construction and scale evaluation studies to establish and verify the scale.

Questionnaire construction and administration

We developed an online questionnaire for use in the United States and United Kingdom using the pool of attitudinal items retained from the item generation stage.

For the stimulus, we presented each participant with four audio recordings of the same short passage (approximately 30 seconds in length), each read by a speaker with a different regional accent of English, and asked them to rate each speaker on all 20 attitudinal items using a 7-point Likert-type scale from low (1) to neutral (4) to high (7). All recordings consisted of the first paragraph of *Comma Gets a Cure* (see Acknowledgements), a passage which uses the Wells Standard Lexical Sets for English [65] to highlight the most differentiable elements of accents.

We selected accents from the dialect regions defined by Labov et al. [66] for the United States and Shackleton [67] for the United Kingdom. The US-based accents in this study were American West and Inland South and the UK-based accents were Received Pronunciation ("RP") and Northwest England. We recorded a speaker from urban Colorado to represent the American West accent, and for the other three accents we used recordings under license from the International Dialects of English Archive ("IDEA"; see Acknowledgements). A full list of the recordings used and speaker demographics is available in the Supplementary Materials.

The IDEA data sources predominately represented white male speakers. As a result of controlling for speaker demographics and audio quality from the available recordings, our speakers all self-identified as white males ranging from 42 to 59 years old. In the sociolinguistic sense, American West (which phonologically is in the spectrum of the "General American" accent) and Received Pronunciation represent standard or "high-prestige" variants within the US and UK, respectively, and Inland South and Northwest England are nonstandard "low-prestige" variants [37,38,68]. The American West and RP speakers used for this study held university degrees and the Inland South and Northwest England speakers did not. The American West and RP speakers were employed in professional teaching occupations and the Northwest England and Inland South speakers were employed in skilled trades. Therefore, their educational and occupational attainment matched the indexical class and status associated with their accents. We presented all participants with all four recordings, regardless of their location.

Prior to being presented with the recordings or giving attitudinal responses, participants each completed a triad test [57] with a lambda-3 balanced incomplete block design [69] for the 11 prestige domain items, resulting in 55 triadic comparisons per participant. In each comparison, participants chose which of the three items was perceived to be least like the others, thereby creating a pair of like items. This could be used to assess whether the perception of the structure of prestige items was consistent beyond the sociolinguistic context of the prestige of regional accents.

We collected a number of demographic variables from participants, to be able to examine any systematic differences in responses. The demographic variables chosen were: country, age, gender, ethnicity, locality size, English proficiency, education, occupation, and income. Each variable and its levels are described in detail in the Supplementary Materials and their distributions within the sample are displayed in **Fig 7** in comparison with those of the subsequent scale evaluation study.

We collected data in May and June 2016 using online surveys implemented on SurveyMonkey and distributed using social media (n = 5 US, 2 UK), the Amazon Mechanical Turk and TurkPrime [70] platforms (n = 148 US), and the Prolific platform (n = 153 UK), for a final sample of 308 (153 US, 155 UK). There were 5 participants (4 US, 1 UK) that completed the triad test but not the attitudinal speech evaluation, so the final sample for the attitudinal data was 303 (149 US, 154 UK). There were otherwise no missing attitudinal or triad data, as we required participants to complete every item in order to receive payment.

Exploratory factor analysis

First, we checked the data for conformity to the assumptions of exploratory factor analysis ("EFA"). Though strict multivariate normality is not required for exploratory or confirmatory methods using categorical models, and violations are allowable under continuous models (i.e. maximum likelihood) if measurement invariance is established [71], we found that the distribution of responses to the attitudinal items was not multivariate normal, with $p \cong 0$ for Mardia's test [72,73], the Henze-Zirkler test [74], and Royston's test [75,76]. We identified multivariate outliers using adjusted chi-square quantile-quantile plots of Mahalanobis distances and removed one participant (from the US sample) with extreme outlier values.

We then assessed the distributions of attitudinal items for approximate univariate normality, as well as for acceptable values of skewness and kurtosis. Following Bulmer [77], absolute values of skewness below 0.5 indicated an approximately symmetric distribution, values between 0.5 and 1.0 were considered moderately skewed, and values above 1.0 were highly skewed. According to the findings of West et al. [78] and Curran et al. [79], issues of bias due to non-normality may result from the analysis of data distributed with absolute skewness values above 2.0 or kurtosis values above 7.0. We found individual variables to be approximately normal and values of skewness (*M* = -0.242, *SD* = 0.407) and kurtosis (*M* = -0.536, *SD* = 0.476) to be within acceptable ranges.

We evaluated linear relationships between items and their factorability by examining inter-item correlations, using the Kaiser-Meyer-Olkin ("KMO") test of sampling adequacy [80], with values greater than or equal to 0.50 considered suitable [81,82,45], and using Bartlett's test of sphericity [83] to test whether the correlation matrix was factorable. We calculated a polychoric correlation matrix because attitudinal items were measured using an ordinal scale [84]. Following Savalei [85], no adjustments were made to zero frequency cells in the bivariate tables. A large proportion of inter-item correlations (73/210, or 34.8%) were above 0.50, indicating the presence of linear relationships. KMO values were well above 0.50 for all variables (overall = 0.946, *M* = 0.935, *SD* = 0.039) and the result of the Bartlett's test was highly significant ($p \cong 0$), together indicating suitable factorability.

Lastly, we evaluated whether our sample sizes were adequate, using the guidelines of having a total sample size of at least $\frac{p(p-1)}{2}$ [86], where p is the number of items or variables, and a subjects-to-variables ratio of at least 10:1 [81] or 20:1 [45]. The sample size for this study (after outlier removal) was 302, which (at $p = 20$) exceeds the suggested minimum of 190, and the subjects-to-variables ratio was 15.1:1, which lies above the recommendation of 10:1 and below 20:1.

We then conducted exploratory factor analysis for the purpose of exploring the structure and dimensionality of the prestige construct. Our analyses used a three-stage robust diagonally weighted least squares estimation technique (weighted least squares, mean and variance adjusted, or "WLSMV") due to its suitability for use on ordinal data with an adequate number of categories [86–88]. We used a conservative oblimin (oblique) factor rotation method to allow for potential intercorrelations between factors, which may be expected in real-world attitudinal data [89].

We eliminated items sequentially, first to remove items that had poor value in discriminating the prestige domain from the other two domains included—solidarity and dynamism—and then to determine the most parsimonious structure within the prestige domain. Items needed to meet all of the following acceptance criteria to be retained: *a*) primary factor loading with an absolute value > 0.32; *b*) cross-loadings with absolute values < 0.32; *c*) gap between primary and cross-loadings > 0.2; and *d*) communality > 0.4 [90]. We re-evaluated the optimal number of factors at each step using the parallel analysis with comparison data method of Ruscio & Roche [91].

Through this process, we obtained the overall factor structure for the attitudinal items across all three domains (**Table 2A**; **Fig 4**), as well as the internal factor structure of the prestige domain items (**Table 2B**; **Fig 1**). Using EFA, items within the prestige domain were partitioned into three factors: *wealthy*, *powerful*, and *high social status* in the first factor, hereafter referred to as "position"; *reputable* and *respected* in the second factor, referred to as "reputation"; and *educated* and *intelligent* in the third factor, referred to as "information." We therefore denote the resulting factor structure as Position-Reputation-Information, or "PRI."

After completing EFA using the attitudinal data, we then repeated the process using the data from the triad test as a second, parallel source of information on the structure of the prestige construct absent the embedded sociolinguistic context. Since the pairings in the triad data are represented as a series of dichotomous observations, we calculated a tetrachoric correlation matrix [92], using a correction of 0.5 for empty bivariate cells (following Savalei [85]) and eigenvector smoothing to ensure the matrix was positive definite. We chose related methods to maximize comparability between the attitudinal and triad data sources. We used a non-robust weighted least squares ("WLS") estimator with standard parallel analysis and identical acceptance criteria to those used for the EFA of the attitudinal data described above.

The inter-item correlations between triad items had 4/55 (7.3%) above 0.50. The overall KMO value was 0.364 (M = 0.368, SD = 0.160), with the lowest individual values being *successful* at 0.075 and *powerful* at 0.157, and the highest being *wealthy* at 0.585. While the result of the Bartlett's test was highly significant ($p \cong 0$), it is also dependent upon sample size, which was reasonably large (n = 308). Taken together, these results suggested that factorability could be poor due to the nature of how the data were represented; specifically, the triadic comparisons generated a matrix with a large amount of "missing" data, as only 3 items in each observation (out of 11 total) had values. The sample size for the triad data was much higher than the suggested minimum of 45 in this case (given the lower number of items), and the subjects-to-variables ratio was 30.8:1, which is above both recommended values. We obtained the internal factor structure for the prestige domain items in the triad data (**Table 2C**) using the EFA methods described. The structure closely resembled the attitudinal results in all respects except that *powerful* was dropped from the position factor due to negative loadings and low communality.

Table 2. Factor loadings and communalities from exploratory factor analysis of attitudinal and triad data. Values are from analyses of: **(A)** all items from attitudinal data over the three domains; **(B)** internal prestige domain items from attitudinal data; and **(C)** internal prestige domain items from triad data. All cases show the items remaining after eliminating prestige domain items that failed to meet acceptance criteria, and after removing *prestigious* from **(B)** and **(C)**. *Note:* Factor loadings with absolute value < 0.20 are suppressed.

(A)

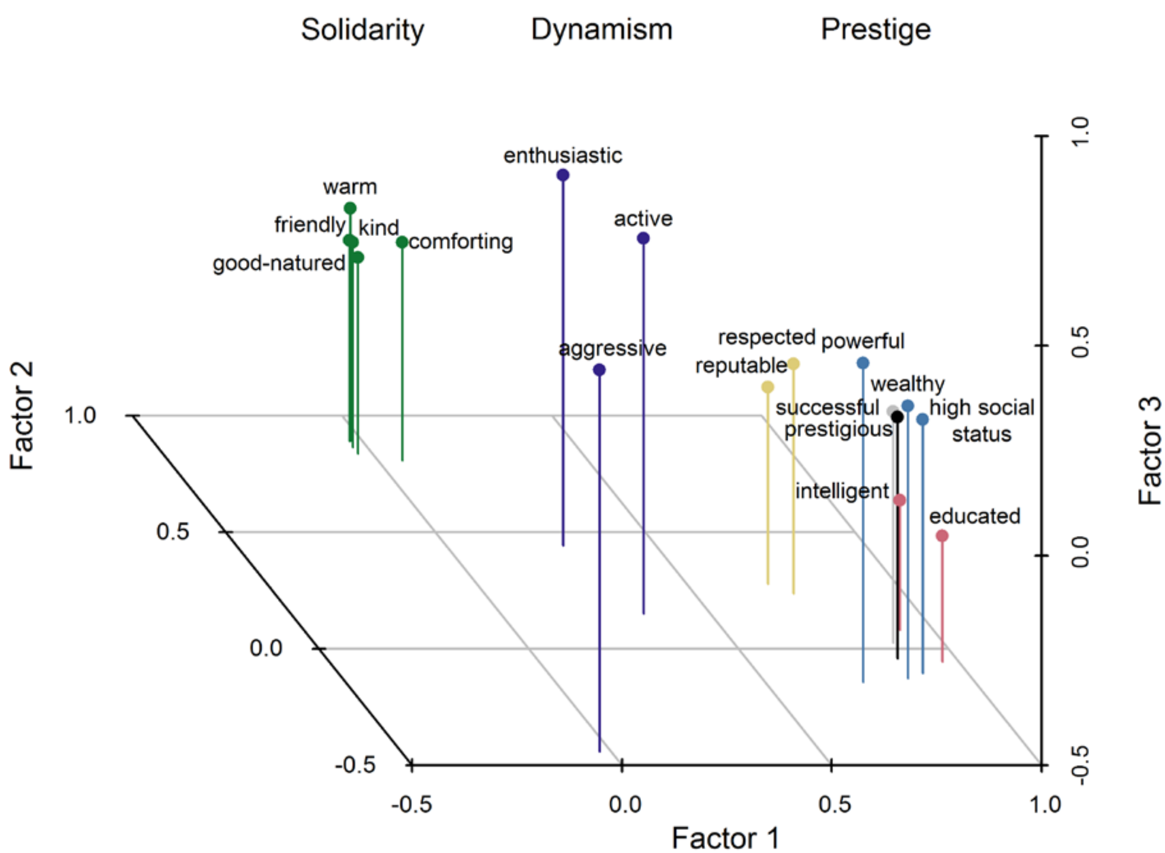
	PRESTIGE	SOLIDARITY	DYNAMISM	communality
educated	0.962	-0.201		0.852
intelligent	0.921			0.806
high social status	0.893			0.862
successful	0.880			0.805
prestigious	0.861			0.781
wealthy	0.847			0.818
respected	0.736	0.237		0.638
powerful	0.734		0.260	0.727
reputable	0.694	0.278		0.573
warm		0.893		0.787
friendly		0.890		0.795
kind		0.865		0.749
good-natured		0.837		0.705
comforting		0.808		0.660
enthusiastic	0.278	0.442	0.383	0.460
aggressive		-0.441	0.409	0.394
active	0.341		0.394	0.364

(B)

	Position	Reputation	Information	communality
wealthy	0.935			0.862
powerful	0.819			0.688
high social status	0.771			0.872
reputable		0.824		0.729
respected	0.238	0.590		0.668
educated			0.893	0.880
intelligent			0.845	0.826

(C)

	Position	Reputation	Information	communality
wealthy	0.776			0.571
high social status	0.657			0.497
reputable		0.820		0.655
respected		0.705		0.532
intelligent			0.962	0.891
educated			0.832	0.749



598

599 **Fig 4. Factor loadings from exploratory factor analysis of attitudinal data.**

600 Visual display of the values in **Table 2A**. Position, reputation, and information items are

601 shown in light blue, gold, and pink, respectively. Other prestige items are shown in

black (*prestigious*, not used in scale) and gray (later dropped from internal prestige structure, shown in **Fig 1**). Solidarity items are in green. Dynamism items are in purple.

Cluster analysis

Following the EFA for both the attitudinal data and the triad data, we also elected to conduct cluster analysis on the items in both data sets to compare results with the EFA findings on the internal structure of the prestige construct. Though the outputs of EFA and cluster analysis are qualitatively similar, the two methods have substantively different goals (dimensionality reduction to latent constructs versus classification to subgroups, respectively) and algorithms. We chose the Partitioning Around Medoids ("PAM") method [93], a type of *k*-medoids algorithm in the *k*-means family, due to its flexibility in accommodating various dissimilarity measures and its robustness against outliers. For the attitudinal data, we used Manhattan distances rather than Euclidean due to their suitability for ordinal data [93]. Visual examination of the Manhattan distance matrix using multidimensional scaling suggested that the attitudinal data were amenable to cluster analysis.

We eliminated items sequentially to remove items with poor discriminant value and to determine the internal prestige structure, retaining items which had a positive silhouette width of at least 0.1 and the removal of which did not substantially improve the overall clustering structure (as measured by average silhouette width of the solution). The silhouette width of an item represents the relative consistency of that item within its cluster. At each step, we used the Duda-Hart test [94] to determine whether more than one cluster was supported and the number of clusters was determined by the highest average silhouette width.

The PAM method resulted in a 2-cluster solution for all attitudinal items (**Fig 5**) and a 3-cluster solution for the internal prestige domain items (**Fig 6A**). The average silhouette width of the 2-cluster solution for all items was 0.428, while the next highest, at 4 clusters, was 0.294. For the internal prestige domain items, the average silhouette width of the 3-cluster solution was 0.282, with 0.300 for 2 clusters. However, the Dunn index, or the ratio of minimum inter-cluster distance to maximum intra-cluster distance [95], was 1.052 for the 3-cluster internal solution and 0.882 for the 2-cluster internal solution, indicating that the 3-cluster solution has better validity. These results support the 3-cluster solution for the internal prestige domain items and this solution matches exactly the PRI structure found through EFA.

Applying the PAM method to the triad data gave similar results, with the highest average silhouette width overall (0.388) found for a 3-cluster solution that matched the PRI structure (**Fig 6B**). However, we reached this solution by eliminating the *hardworking* and *ambitious* items based on information from the EFA showing their poor fit within the prestige domain. The triadic comparisons included only prestige domain items so, within the context of the triad data alone, this information about the ability to discriminate from other domains would be unavailable. Additionally, the Dunn index suggested better support for this 3-cluster solution (1.112) than for a 4-cluster solution that included *hardworking* and *ambitious* (1.051). These results are consistent with what we found from the attitudinal data and replicate the PRI structure in the best-fitting solution.

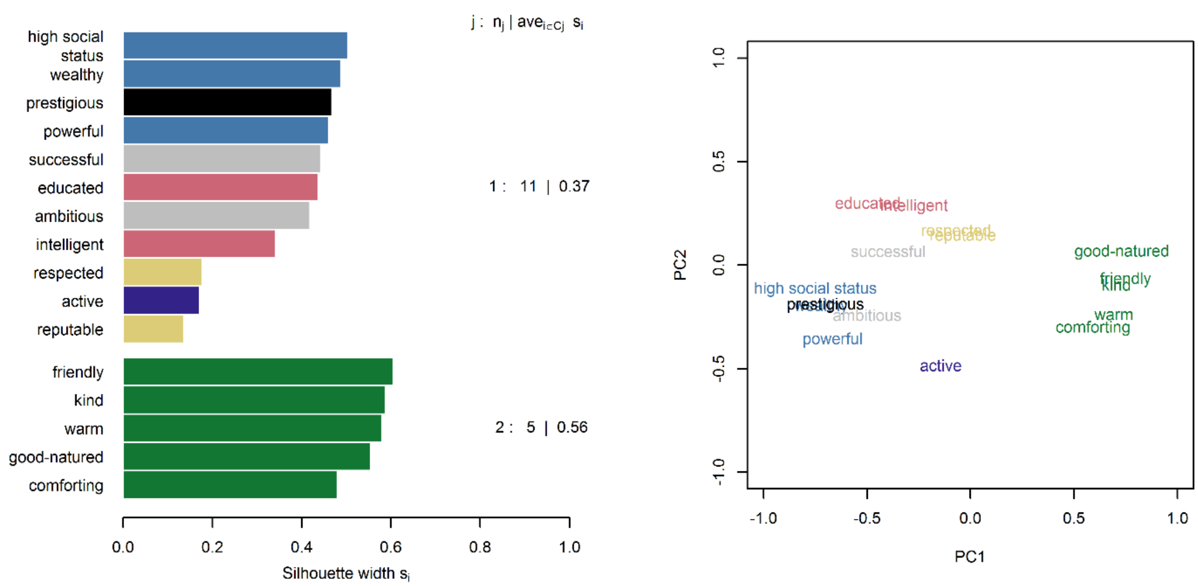


Fig 5. Silhouette and multidimensional scaling plots of the clustering of prestige, solidarity, and dynamism attitudinal items. Position, reputation, and information items are shown in light blue, gold, and light red, respectively. Other prestige items are

shown in black (*prestigious*, not used in scale) and grey (later dropped from internal prestige structure, shown in **Fig 6A**). Solidarity items are in green. The remaining dynamism item (*active*) is in purple. To the right of each cluster is the number of items in that cluster and its average silhouette width. Silhouette width values represent the relative consistency of each item within its cluster.

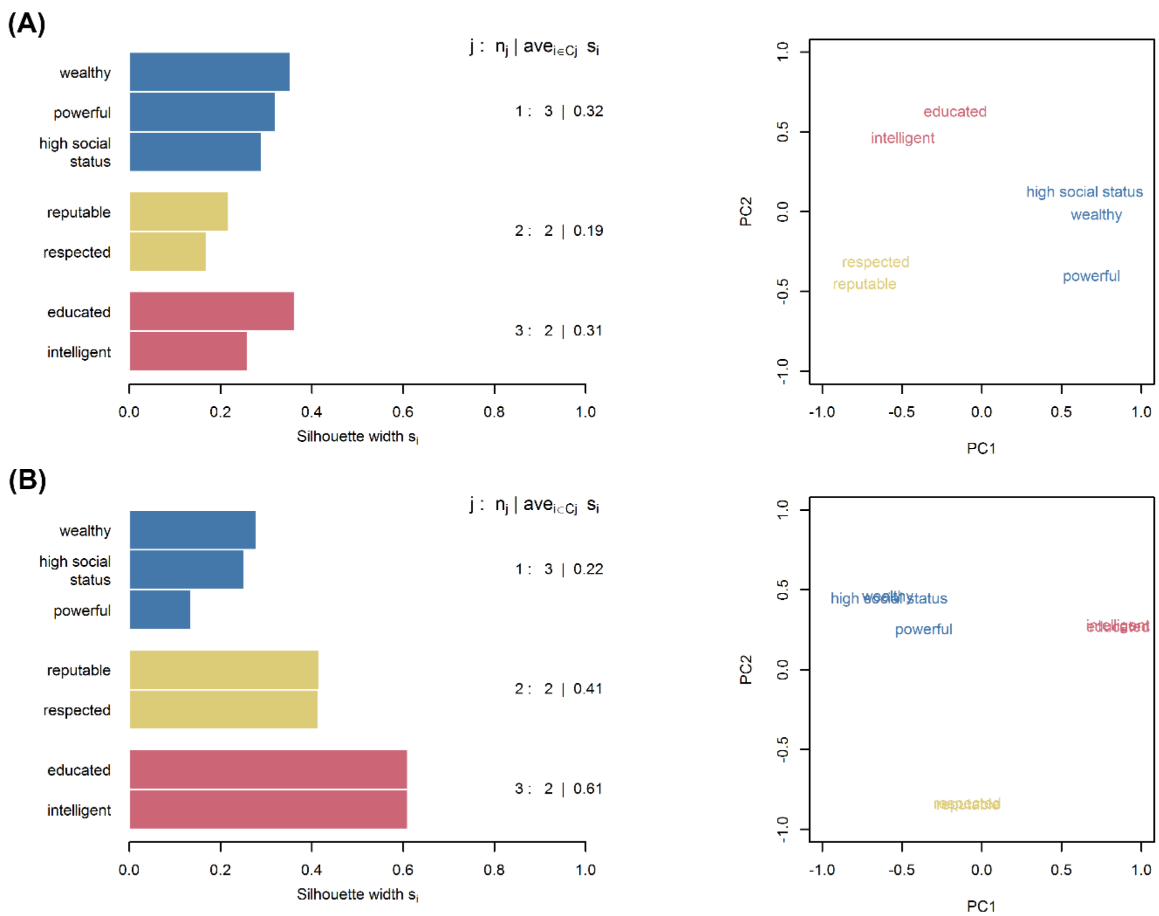


Fig 6. Silhouette plots and multidimensional scaling plots of the clustering of prestige domain items. Plots depict clustering of: **(A)** internal prestige domain

attitudinal items, after eliminating items that failed to meet acceptance criteria; and **(B)** prestige domain triad items, after eliminating items that failed to meet acceptance criteria as well as *hardworking* and *ambitious* (see text). Position, reputation, and information items are shown in light blue, gold, and light red, respectively. To the right of each cluster is the number of items in that cluster and its average silhouette width. Silhouette width values represent the relative consistency of each item within its cluster. Multidimensional scaling was done using Manhattan distances for attitudinal items and tetrachoric correlations for triad items.

Study 2: Scale evaluation

Item generation

We used the full set of items generated for the previous scale construction study (**Table 1**) for evaluation and validation of the scale. We selected three additional prestige items (*talented*, *driven*, and *skilled*) from those generated by the free listing exercise to explore whether the inclusion of additional terms would have any effect on the PRI structure or provide additional explanatory power. As a number of the existing terms could be interpreted as measures of "ascribed" prestige (i.e. traits that are largely assigned or fixed based on the circumstances of one's birth), we chose these terms as representative of the concept of "achieved" prestige (i.e. traits that can be earned or acquired) [34,96].

We also reverse-scored three items (*intelligent-unintelligent*, *ambitious-unambitious*, and *kind-unkind*) to reduce potential bias in responses [97], selected intentionally to avoid potentially ambiguous reversals. However, during exploratory analyses, we found that the distributions of responses to the reversed items were significantly skewed toward higher values than for the same items in the scale construction study. This suggests that participants were less likely to agree with a negative assessment of a speaker (i.e. *unintelligent*) than they were to disagree with its opposite positive assessment (*intelligent*). These differences caused issues with the consistency of responses and negatively affected model fit, similar to the problems seen later with reversed items in the Cheng et al. [26] scale (see Criterion validity) but to a lesser degree. Due to these issues, we do not recommend reversal for future studies using attitudinal items scored on a Likert-type scale (cf. [98]).

Questionnaire construction and administration

In the online questionnaire for the scale evaluation study, we presented each participant with 10 audio recordings of the same passage used in the scale construction study: the first paragraph of *Comma Gets a Cure*. Each recording used a speaker with a different regional accent of English, and we asked participants to rate each speaker on all 23 attitudinal items (**Table 1**, plus *talented*, *driven*, and *skilled* under prestige) using a 7-point Likert-type scale from strongly disagree (1) to strongly agree (7).

We presented participants in the US with 8 US-based accents and 2 UK-based accents, while participants in the UK were presented with 8 UK-based accents and 2 US-based accents, for a total of 16 different accents across the entire sample, 4 of which were cross-tested in both countries (**Table 3**). The recordings for the 4 cross-tested accents were identical to those used in the scale construction study. All recordings were used under license from IDEA (see Acknowledgements) except for American West (Urban) and Wales, which we recruited from local contacts and recorded.

Table 3. Regional accents and the participants to which they were presented in the scale evaluation study.

United States	All Participants	United Kingdom
American West (Rural)	American West (Urban)	Southwest England
Midland	American Inland South (Blue-Collar)	Southeast England
Inland North	Received Pronunciation	Yorkshire
American Inland South (White-Collar)	Northwest England	Scotland
Mid-Atlantic		Ireland
New York City		Wales

As in the scale construction study, we selected speakers for consistency from the recordings available. All speakers self-identified as white men and ranged in age from

31 to 59 years. Speakers varied in their level of education, occupation, and settlement size during childhood. The speaker from Wales was 45 years old at the time of recording, held an advanced degree, and was employed in an academic profession. The demographics of the American West (Urban) speaker are given in the methods for the scale construction study and all speaker demographics are available in the Supplementary Materials.

We collected data in June 2016 using online surveys implemented on SurveyMonkey and distributed using Amazon Mechanical Turk and TurkPrime [70] (*n* = 151 US) and Prolific (*n* = 144 UK), for a sample size of 295. We excluded participants from the prior scale construction study to ensure an independent sample. The results did not contain any missing data for attitudinal items.

Demographic comparisons

Demographic characteristics of the scale evaluation sample were similar to the scale construction sample (**Fig 7**). Permutation tests of independence [99], adjusted for multiple comparisons to control for the false discovery rate, confirmed that significant differences were present only in the distributions of the age ($p < 0.001$) and occupation ($p < 0.001$) variables between the two studies, as a result of a larger proportion of relatively younger students in the scale evaluation sample. Given the similarity across all other variables, we considered this to be a relatively minor issue, and one that could be checked analytically by examining measurement invariance (see Confirmatory factor analysis).

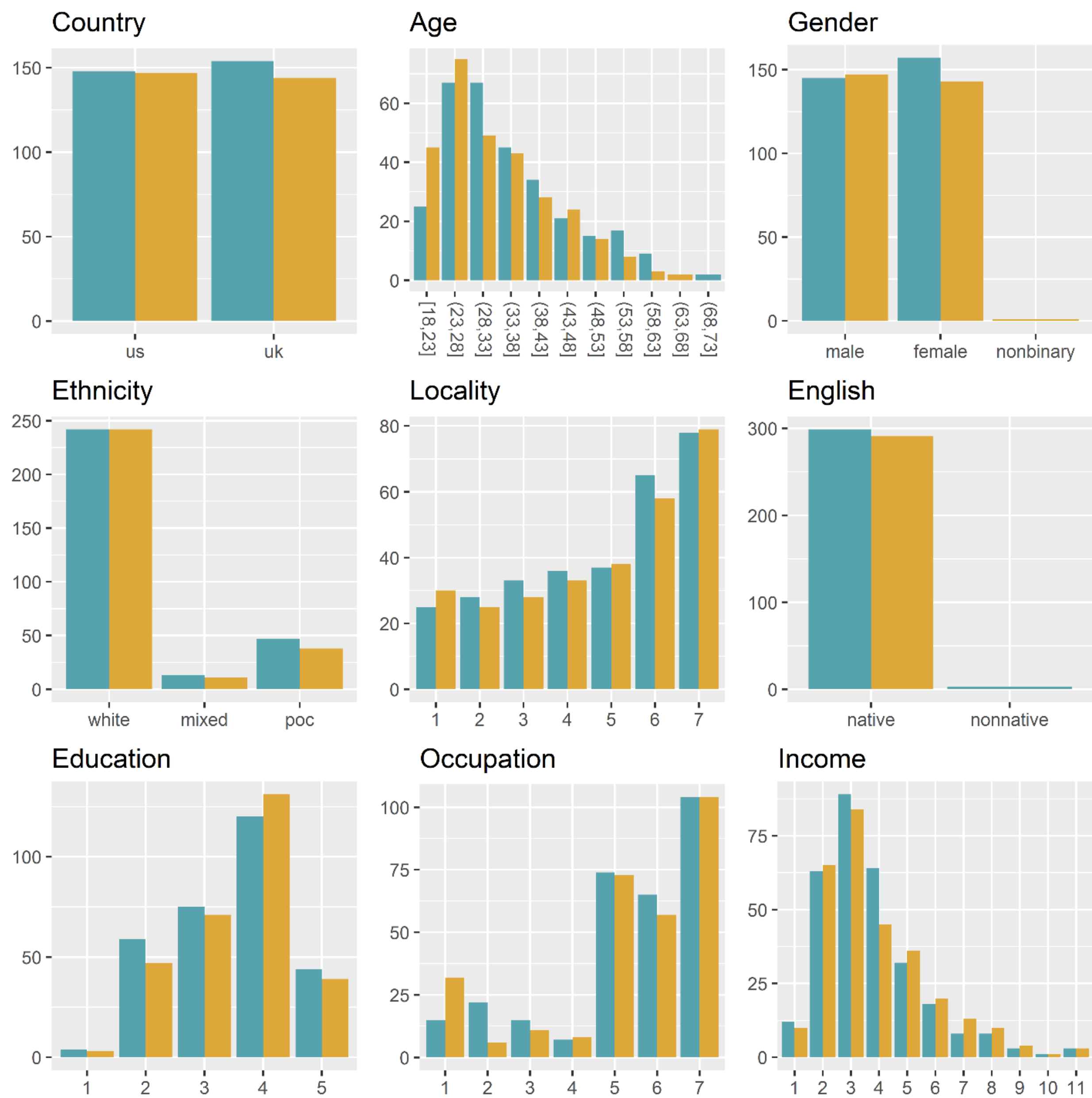


Fig 7. Demographic comparisons of the samples from the scale construction and scale evaluation studies. Vertical axes display counts of participants for each variable. Teal bars represent participants in the first (scale construction) study and orange bars represent participants in the second (scale evaluation) study. Numerical levels for categorical variables are defined in the Supplementary Materials.

Exploratory factor analysis

Following checks of assumptions, outliers, item relationships, item factorability, and sample size, we conducted EFA on the scale evaluation data using methods and criteria identical to those used in the scale construction study, to address the question of whether the items generated adequately represented the breadth and structure of the individual prestige concept. The items that were previously eliminated in the EFA of the scale construction study were eliminated again in the process of conducting this EFA, due to violations of acceptance criteria. All three additional "achieved" prestige items—*talented*, *driven*, and *skilled*—were also eliminated, particularly because of high cross-loadings or primary loadings on other factors. We therefore made no changes to the structure of the model or the items included and found the scope of the existing PRI model to be adequate for use in CFA.

Confirmatory factor analysis

The distribution of responses to the attitudinal items was not multivariate normal ($p \cong 0$ for all tests). We removed four participants with six extreme outlier values (all from the US sample) as a result of examining Mahalanobis distances, leaving a final sample size of 291. The individual variables were approximately normal, and values of skewness (M = -0.209, SD = 0.369) and kurtosis (M = -0.617, SD = 0.279) were within acceptable ranges.

We then performed measurement invariance testing [100], to ensure that the relationships between indicators and latent variables within the prestige construct were consistent across participant demographic groups by country, age, gender, ethnicity, locality size, educational attainment, occupation, and income (see Supplementary Materials for details on demographic variables). The sample contained an insufficient number of non-native English speakers to test for invariance by native English proficiency; therefore, we excluded this variable. We tested five increasingly constrained models in sequence: configural invariance (Model 1), metric or "weak" invariance (Model 2), scalar or "strong" invariance (Model 3), residual or "strict" invariance (Model 4), and residual invariance with constrained means (Model 5). We established configural invariance using the permutation method proposed by Jorgensen et al. [101]. We looked at changes in two noncentrality-based fit indices, the

Comparative Fit Index (“CFI”) and the Root Mean Square Error of Approximation (“RMSEA”), to evaluate the relative fit of each successive nested model, with ΔCFI values less than or equal to 0.010 and ΔRMSEA values less than or equal to 0.015 indicating invariance [102]. Fulfilment of scalar invariance was considered sufficient to proceed with confirmatory factor analysis [100].

Measurement invariance of the model was upheld across the demographic variables of country, age, gender, occupation, and income. We found metric non-invariance by locality size (ΔCFI = 0.011, ΔRMSEA = 0.030), and ethnicity and educational attainment were borderline metric non-invariant (ΔCFI = 0.007, ΔRMSEA = 0.022; and ΔCFI = 0.009, ΔRMSEA = 0.026, respectively). Given these results, we defined a complex survey design which re-fit the model using pseudo-maximum likelihood and provided adjusted point and variance estimates [103,104]. In this design, the potentially non-invariant demographic variables were incorporated as weighted sampling strata using weights approximated from US [105–107] and UK census data [108–110].

We then performed confirmatory factor analysis (“CFA”) to assess the fit of this model to the scale evaluation data. As the WLSMV estimation method used in previous analyses could not be applied to a complex survey design, we used maximum likelihood with robust standard errors (mean and variance adjusted using the Satterthwaite approach [111], “MLMVS”) for the CFA based on the complex survey design. Equivalent results should be obtained from either method, as the two perform comparably for 7-point ordinal data [87].

We assessed the goodness of fit of confirmatory models with and without the complex survey design using two incremental fit indices (the CFI, as above, and the Tucker-Lewis Index, or “TLI”) and two absolute fit indices (the RMSEA, as above, and the Standardized Root Mean Square Residual, or “SRMR”). We drew cutoff criteria from Hu & Bentler [112] and adjusted them to the recommendations of Yu [113] as follows: CFI > 0.95; TLI > 0.96; RMSEA < 0.05; and SRMR < 0.07. We obtained parameter estimates using both robust maximum likelihood and robust weighted least squares methods, and compared fit indices for the models using MLMVS estimation, MLMVS estimation with a complex survey design, and WLSMV estimation (**Table 4**).

Table 4. Goodness of fit indices for the Position-Reputation-Information scale model. Robust indicators were estimated using: maximum likelihood methods, maximum likelihood methods with fit adjustments from a complex survey design to account for demographic non-invariance, and weighted least squares methods. Bounds for 90% confidence intervals are provided for RMSEA.

Model	CFI	TLI	RMSEA [90% CI]	SRMR
MLMVS	0.948	0.983	0.056 [0.048, 0.064]	0.020
MLMVS with Complex Survey Design	0.959	0.983	0.031 [0.026, 0.036]	0.023
WLSMV	0.994	0.989	0.094 [0.085, 0.104]	0.022

All three models shared the identical PRI structure and had comparable fit indices. We selected the model using MLMVS with adjustments from the complex survey design as the preferred model (**Fig 2**) because it fulfilled the cutoff criteria for all fit indices and properly incorporated information on all potentially non-invariant demographic variables.

Scale validity and reliability

Content validity

Content validity is the assessment of whether the scale adequately represents the extent of the domain of interest [114]. As content validity is essentially a qualitative judgment rather than a statistical one [56], we worked to establish and report the content validity of the PRI scale through the methods used to generate the items and those used to construct and verify the scale.

As mentioned previously, items were generated in part by participants through an inductive, endogenous process in the free listing task, which produced a broad but consistent sample of items. We supplemented this with more traditional deductive sampling of terms from previous literature and a pilot study. Rather than consulting external subject matter experts (cf. [115]), we valued more highly the validity of judgments by the study participants in generating and associating items.

Additionally, we included three "achieved" prestige items (*talented*, *driven*, and *skilled*), drawn from the free listed terms, to confirm the content validity of the model being tested. These items (and all of the same items dropped previously from the scale construction study) were dropped due to failure to meet the acceptance criteria, which lends support to the validity of the PRI scale and the sufficiency of its domain breadth.

Finally, we considered that the relative lack of demographic diversity among free listing participants in the initial scale construction study (compared to that of our other samples) could have negatively impacted the breadth of items generated and hence the content validity of the scale. However, we found no specific evidence to suggest this was the case, aside from potential issues of measurement non-invariance (see

Confirmatory factor analysis) which could have occurred regardless. We therefore do not consider this to have been a point of concern for the present study but would recommend future studies endeavor to recruit a maximally diverse and representative sample from the population of interest for item generation.

Construct validity

Construct validity is the property that the scale measures what it is intended to measure, which is generally confirmed by showing correlations among elements expected to be similar ("convergent validity") and a lack of correlation among elements expected to be dissimilar ("discriminant validity"). The construct validity of the PRI scale was established by examining the convergent validity of scale items, the discriminant validity between PRI subscales (position, reputation, and information), and the discriminant validity between the prestige scale and the other two domains included in the data (solidarity and dynamism).

We assessed convergent validity within the scale by examining the polychoric correlation matrices of the scale items in both studies and the average variance explained ("AVE") of the scale items in the scale evaluation study. Following common practice, correlation coefficients between 0.10 and 0.30 were considered small, between 0.30 and 0.50 were moderate, and greater than 0.50 were large [116]. AVE values greater than 0.50 were deemed acceptable, as they indicate sufficient variance attributed to the construct as opposed to measurement error [117].

We found polychoric correlations (ϱ) between all PRI scale items to be high (*M* = 0.631, *SD* = 0.094) and correlations were higher between items within the same subscale than between items in different subscales (**Table 5**). The AVE values for each of the three subscales—position, reputation, and information—were 0.675, 0.630, and 0.699, respectively; all were above the criterion of 0.50, supporting convergent validity.

Table 5. Polychoric correlations between Position-Reputation-Information scale items. Mean (and standard deviation) polychoric correlations between Position-Reputation-Information items within the same subscale are shown on the main diagonal and those for items between different factors are shown below the main diagonal. Values were calculated using the combined scale construction and scale evaluation data sets. Correlations within 2-item factors have no mean or standard deviation because they consist of only one measurement.

	Position	Reputation	Information
Position	0.733 (0.097)		
Reputation	0.580 (0.043)	0.707	
Information	0.625 (0.118)	0.594 (0.054)	0.733

The discriminant validity of constructs, which naturally opposes convergent validity, was assessed using the heterotrait-monotrait ratio of correlations criterion (“HTMT”), a method developed to avoid the potential issues of other indices [118]. For this criterion, lower values indicate greater discriminant validity. HTMT values between the prestige PRI subscales and the other two constructs—solidarity and dynamism—were all below the cutoff of 0.85 advised by Voorhees et al. [119], verifying discriminant validity of the prestige construct (**Table 6**). Similarly, HTMT values showed good discriminant validity between the three PRI subscales. This shows that the three PRI subscales, along with showing good convergent validity (as their items are all measuring elements of the same prestige construct), also exhibit substantial discriminant validity from other constructs and from one another. We consider these results to be support for the PRI scale’s overall construct validity and simple structure.

Table 6. Heterotrait-monotrait ratio of correlations between items from Position-Reputation-Information subscales and solidarity and dynamism constructs. HTMT values between each Position-Reputation-Information subscale and the solidarity and dynamism constructs are shown below the main diagonal. Lower HTMT values indicate greater discriminant validity. Values were calculated using the scale evaluation data set.

		PRESTIGE			SOLIDARITY
		Position	***Reputation***	***Information***	
PRESTIGE	***Position***				
	Reputation	0.818			
	Information	0.841	0.835		
SOLIDARITY		0.086	0.442	0.246	
DYNAMISM		0.727	0.773	0.735	0.670

Criterion validity

The criterion validity of a scale relates to its ability to be used as a measurement tool for the construct of interest, either assessed concurrently with a direct measure of that construct, in comparison with other available tests, or as a predictive indicator of independent or future outcomes. Predictive validity could not be assessed in this instance, as we did not have any future measurements or any independent prestige-related traits that were not already used in scale construction and evaluation, so we assessed the concurrent criterion validity of the scale through the other two avenues.

We first compared each item's polychoric correlation with the *prestigious* item. The *prestigious* item was included in the surveys but excluded from the scale, and was used as a direct representative of the general construct of prestige that we intended to measure. In the scale evaluation data set, polychoric correlations between scale items and the *prestigious* item were high overall (M = 0.678, SD = 0.104), as were mean correlations with *prestigious* within each of the PRI factors (position: M = 0.748, SD = 0.096; reputation: M = 0.626, SD = 0.026; information: M = 0.627, SD = 0.143). Estimated factor scores for each PRI factor (using the Empirical Bayes Modal approach [120] for ordinal variables and the MLMVS model with adjustments from the complex survey design) were even more highly correlated with *prestigious* than the raw item scores (PRI individual prestige: ϱ = 0.815; position: ϱ = 0.844; reputation: ϱ = 0.764; information: ϱ = 0.745).

Secondly, to compare with another test of prestige, we asked a new set of participants (n = 91 US, 53 UK; again recruited through Amazon Mechanical Turk and Prolific) to rate two new speakers (having the Inland South and Received Pronunciation accents) using the present scale alongside the prestige-dominance scale of Cheng et al. (as detailed in the Electronic Supplementary Material of [26]). We modified the text of the items in the Cheng et al. scale (from "members of your/the group" to "people") to better fit the context of our study. We removed outliers from the data and, using the same methods as above (with WLSMV estimation for the Cheng et al. scale data as the previous estimation method was not specified [26]), calculated factor scores for the PRI subscales, the solidarity and dynamism dimensions, and the prestige and dominance factors from the Cheng et al. prestige-dominance scale. We calculated polychoric correlations to examine the level of agreement between these measures.

In this additional comparative data set, we found substantial correlations between factor scores of the PRI scale and the prestige factor of the Cheng et al. scale (PRI overall: $\varrho = 0.850$, position: $\varrho = 0.805$, reputation: $\varrho = 0.861$, information: $\varrho = 0.828$) and, in general, we found that the individual prestige items of each scale were correlated ($M = 0.567$, $SD = 0.221$). However, one item in particular from the Cheng et al. scale (item 17: “Other people do NOT enjoy hanging out with him”) was relatively uncorrelated with PRI items and with the other Cheng et al. prestige items. Notably, this is one of the three reversed items in the Cheng et al. prestige factor, the other two of which (items 2 and 6: “People do NOT want to be like him” and “People do NOT value his opinion”) had only moderate correlations with PRI items and other Cheng et al. prestige items. The removal of all three reversed items had little effect on correlations between the Cheng et al. prestige factor and the PRI subscales (PRI overall: $\varrho = 0.856$, position: $\varrho = 0.810$, reputation: $\varrho = 0.867$, information: $\varrho = 0.832$) but improved the mean correlation between individual items ($M = 0.690$, $SD = 0.066$).

The reversed items contributed to the poor fit of the Cheng et al. scale overall in this data set (CFI = 0.875, TLI = 0.856, RMSEA = 0.229 [90% CI: 0.219, 0.238], SRMR = 0.154; using WLSMV estimation). The model fit improved with the removal of all reversed items (CFI = 0.973, TLI = 0.966, RMSEA = 0.151 [90% CI: 0.137, 0.165], SRMR = 0.083), but remained unacceptable under criteria for the two absolute fit indices, RMSEA and SRMR. We found the fit of the PRI scale using the same data and estimation method (WLSMV) met the cutoffs for all indices except RMSEA (CFI = 0.998, TLI = 0.995, RMSEA = 0.106 [90% CI: 0.075, 0.139], SRMR = 0.019). Notably, polychoric correlations between—first—the factor scores for dominance in the Cheng et al. scale (reversed items removed) and—second—the Cheng et al. prestige factor scores, the *prestigious* item, and the PRI factor scores, were all moderate to high (Cheng et al. prestige: $\varrho = 0.449$, *prestigious*: $\varrho = 0.561$, PRI prestige overall: $\varrho = 0.533$, position: $\varrho = 0.569$, reputation: $\varrho = 0.489$, information: $\varrho = 0.501$), which may indicate issues with the validity of the dominance construct.

Interrater reliability

In these studies, we did not expect participants to rate each speaker identically for each item, nor is such agreement required to obtain a reliable scale of individual prestige. As mentioned in the Introduction, prior work has shown that different demographic groups will evaluate accents differently. By testing and adjusting the fit of the confirmatory model, we already incorporated information on patterns of variation in item ratings, both by individual and between demographic groups. Our results showed that participants displayed a consistent understanding of the overall prestige construct regardless of disagreements about particular speakers. This being said, measures of interrater reliability can be obtained and so we provide them here for completeness.

We calculated Krippendorff's alpha coefficient [121] using ordinal weights, as well as the intraclass correlation coefficient ("ICC," specifically ICC(C,1) of McGraw & Wong [122]). The level of Krippendorff's alpha indicating agreement was 0.8, with values between 0.667 and 0.800 allowing for "tentative conclusions" [121]. For the ICC, values less than 0.40 were considered to be poor, between 0.40 and 0.60 were fair, between 0.60 and 0.75 were good, and greater than 0.75 were excellent [123]. The reliability values of Krippendorff's alpha obtained for the scale construction and scale evaluation data sets were 0.414 and 0.383, respectively. ICC values for the two data sets were 0.473 [95% CI: 0.359, 0.625] and 0.459 [95% CI: 0.346, 0.612], using only the ratings of speakers that were cross-tested in both countries.

Internal consistency

Lastly, the internal consistency of a scale measures the similarity of results across scale items. We examined this by calculating Cronbach's alpha [124] as well as three variations of the omega coefficient (Raykov, Bentler, and McDonald, as described in [125]). The criterion used for acceptable values of internal consistency measures, given that this study is basic research for the purpose of developing a scale, was 0.80 [81,126].

Using the fitted MLMVS model with adjustments from the complex survey design, internal consistency measures were well above the cutoff for the overall scale, and above or slightly below it for the three PRI latent factors (**Table 7**). Analyses showed that these values would only decrease we removed any individual scale item, suggesting that they are all vital to the structure of the scale.

Table 7. Internal consistency measures for the Position-Reputation-Information scale and its subscales.

	Cronbach's alpha	Omega		
PRESTIGE	0.892	0.918		
		Raykov	**Bentler**	**McDonald**
Position	0.844	0.858	0.858	0.859
Reputation	0.772	0.773	0.773	0.773
Information	0.794	0.818	0.818	0.818

Acknowledgements

We would like to thank Russell Gray for his advisement of this project. We would also like to thank Jerry Vaske and Jeffrey Snodgrass for their valuable feedback, Mark Prince for help with structural equation modeling methods, and Cory Holland for advice on speech recording and accent evaluation.

The recordings used in this project—except for American West (Urban) and Wales, recorded by the authors—are used by special permission of the International Dialects of English Archive, online at http://www.dialectsarchive.com, and we are grateful to IDEA for the licensed use of these recordings. *Comma Gets a Cure* is copyright 2000 Douglas N. Honorof, Jill McCullough & Barbara Somerville, text available online at: http://www.dialectsarchive.com/comma-gets-a-cure.

The color palettes used in figures are derived from a technical note by Paul Tol [127] and are optimized for color-blind readers.

References

1. Asch SE. The doctrine of suggestion, prestige and imitation in social psychology. Psychological Review. 1948;55: 250–276. doi:10.1037/h0057270
2. Wegener B. Concepts and measurement of prestige. Annual Review of Sociology. 1992;18: 253–280. doi:10.1146/annurev.so.18.080192.001345
3. Barkow JH. Prestige and culture: A biosocial interpretation. Current Anthropology. 1975;16: 553–572. doi:10.1086/201619
4. Harbaugh WT. The prestige motive for making charitable transfers. The American Economic Review. 1998;88: 277–282.
5. Boyd R, Richerson PJ. Culture and the evolutionary process. University of Chicago Press; 1985.
6. Henrich J. Cultural transmission and the diffusion of innovations: Adoption dynamics indicate that biased cultural transmission is the predominate force in behavioral change. American Anthropologist. 2001;103: 992–1013. doi:10.1525/aa.2001.103.4.992
7. Henrich J, Boyd R. On modeling cognition and culture. Journal of Cognition and Culture. 2002;2: 87–112. doi:10.1163/156853702320281836
8. Richerson PJ, Boyd R. Not by genes alone: How culture transformed human evolution. University of Chicago Press; 2005.
9. Guppy N, Goyder JC. Consensus on occupational prestige: A reassessment of the evidence. Social Forces. 1984;62: 709–725. doi:10.1093/sf/62.3.709
10. Goyder J. The dynamics of occupational prestige: 1975-2000. Canadian Review of Sociology and Anthropology. 2005;42: 1–23. doi:10.1111/j.1755-618X.2005.tb00788.x
11. Crawley D. Gender and perceptions of occupational prestige. SAGE Open. 2014;4: 1–11. doi:10.1177/2158244013518923
12. Kitchin R. Big data, new epistemologies and paradigm shifts. Big Data & Society. 2014;1: 2053951714528481. doi:10.1177/2053951714528481

13. Mazzocchi F. Could big data be the end of theory in science? EMBO reports. 2015;16: e201541001. doi:10.15252/embr.201541001

14. Mael F, Ashforth BE. Alumni and their alma mater: A partial test of the reformulated model of organizational identification. Journal of Organizational Behavior. 1992;13: 103–123. doi:10.1002/job.4030130202

15. Smidts A, Pruyn ATH, van Riel CBM. The impact of employee communication and perceived external prestige on organizational identification. Academy of Management Journal. 2001;44: 1051–1062. doi:10.2307/3069448

16. Deeter-Schmelz DR, Moore JN, Goebel DJ. Prestige clothing shopping by consumers: a confirmatory assessment and refinement of the PRECON scale with managerial implications. Journal of Marketing Theory and Practice. 2000;8: 43–58. doi:10.1080/10696679.2000.11501879

17. Vigneron F, Johnson LW. Measuring perceptions of brand luxury. Journal of Brand Management. 2004;11: 484–506. doi:10.1057/palgrave.bm.2540194

18. Duncan OD. A socioeconomic index for all occupations. In: Reiss AJ, Duncan OD, Hatt PK, North CC, editors. Occupations and social status. New York: Free Press; 1961. pp. 109–138.

19. Nakao K, Treas J. The 1989 socioeconomic index of occupations: Construction from the 1989 occupational prestige scores. Chicago: National Opinion Research Center; 1992.

20. Ganzeboom HB, De Graaf PM, Treiman DJ. A standard international socio-economic index of occupational status. Social Science Research. 1992;21: 1–56. doi:10.1016/0049-089X(92)90017-B

21. Treiman DJ. Occupational prestige in comparative perspective. New York: Academic Press; 1977.

22. Kaufman HF. Defining prestige rank in a rural community. Sociometry. 1945;8: 199–207. doi:10.2307/2785239

23. Steenkamp JBE, Batra R, Alden DL. How perceived brand globalness creates brand value. Journal of International Business Studies. 2003;34: 53–65. doi:10.1057/palgrave.jibs.8400002

24. Gusfield JR, Schwartz M. The meanings of occupational prestige: Reconsideration of the NORC scale. American Sociological Review. 1963;28: 265–271. doi:10.2307/2090613

25. Blaikie NW. The meaning and measurement of occupational prestige. Australian & New Zealand Journal of Sociology. 1977;13. doi:10.1177/144078337701300202

26. Cheng JT, Tracy JL, Henrich J. Pride, personality, and the evolutionary foundations of human social status. Evolution and Human Behavior. 2010;31: 334–347. doi:10.1016/j.evolhumbehav.2010.02.004

27. Henrich J, Gil-White FJ. The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. Evolution and human behavior. 2001;22: 165–196. doi:10.1016/S1090-5138(00)00071-4

28. Goldthorpe JH, Hope K. Occupational grading and occupational prestige. In: Hope K, editor. The analysis of social mobility: Methods and approaches. Oxford: Clarendon Press; 1972. pp. 19–80.

29. Seligson MA. Prestige among peasants: A multidimensional analysis of preference data. American Journal of Sociology. 1977;83: 632–652. doi:10.1086/226597

30. Reyes-García V, Molina JL, Broesch J, Calvet L, Huanca T, Saus J, et al. Do the aged and knowledgeable men enjoy more prestige? A test of predictions from the prestige-bias model of cultural transmission. Evolution and Human Behavior. 2008;29: 275–281. doi:10.1016/j.evolhumbehav.2008.02.002

31. von Rueden C, Gurven M, Kaplan H. The multiple dimensions of male social status in an Amazonian society. Evolution and Human Behavior. 2008;29: 402–415. doi:10.1016/j.evolhumbehav.2008.05.001

32. Davis K, Moore WE. Some principles of stratification. American Sociological Review. 1945;10: 242–249. doi:10.2307/2085643

33. Inkeles A, Rossi PH. National comparisons of occupational prestige. American Journal of Sociology. 1956;61: 329–339. doi:10.2307/2773534

34. Svalastoga K. Prestige, class, and mobility. Scandinavian University Books; 1959.

35. D'Aveni RA. Top managerial prestige and organizational bankruptcy. Organization Science. 1990;1: 121–142. doi:10.1287/orsc.1.2.121

36. Vigneron F, Johnson LW. A review and a conceptual framework of prestige-seeking consumer behavior. Academy of Marketing Science Review. 1999;1: 1–15.

37. Giles H. Evaluative Reactions to Accents. Educational Review. 1970;22: 211–227. doi:10.1080/0013191700220301

38. Bishop H, Coupland N, Garrett P. Conceptual accent evaluation: Thirty years of accent prejudice in the UK. Acta Linguistica Hafniensia. 2005;37: 131–154. doi:10.1080/03740463.2005.10416087

39. Coupland N, Bishop H. Ideologised values for British accents. Journal of Sociolinguistics. 2007;11: 74–93.

40. Fuertes JN, Gottdiener WH, Martin H, Gilbert TC, Giles H. A meta-analysis of the effects of speakers' accents on interpersonal evaluations: Effects of speakers' accents. European Journal of Social Psychology. 2012;42: 120–133. doi:10.1002/ejsp.862

41. Cronk L. The application of animal signaling theory to human phenomena: Some thoughts and clarifications. Social Science Information. 2005;44: 603–620. doi:10.1177/0539018405058203

42. Kroch AS. Toward a theory of social dialect variation. Language in Society. 1978;7: 17–36. doi:10.1017/S0047404500005315

43. Kahane H. A typology of the prestige language. Language. 1986;62: 495–508. doi:10.2307/415474

44. Labov W. The social stratification of English in New York City. Washington, D.C.: Center for Applied Linguistics; 1966.

45. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate data analysis. 10th ed. Upper Saddle River, NJ: Prentice Hall; 2010.

46. Weber M. Wirtschaft und gesellschaft: Grundriss der verstehenden soziologie. Weber M, editor. Tübingen: Mohr Siebeck; 1922.

47. Weber M. From Max Weber: Essays in sociology. Gerth HH, Mills CW, editors. New York: Oxford University Press; 1946.

48. Rhoads JK. Critical issues in social theory. University Park, PA: Penn State Press; 2010.

49. Henrich J, Gil-White FJ. The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. Evolution and Human Behavior. 2001;22: 165–196. doi:10.1016/S1090-5138(00)00071-4

50. Shenkar O, Yuchtman-Yaar E. Reputation, image, prestige, and goodwill: An interdisciplinary approach to organizational standing. Human Relations. 1997;50: 1361–1381. doi:10.1177/001872679705001102

51. Bajema CJ. A note on the interrelations among intellectual ability, educational attainment, and occupational achievement: A follow-up study of a male Kalamazoo public school population. Sociology of Education. 1968;41: 317–319. doi:10.2307/2111879

52. Sewell WH, Haller AO, Portes A. The educational and early occupational attainment process. American Sociological Review. 1969;34: 82–92. doi:10.2307/2092789

53. Cheng H, Furnham A. Childhood cognitive ability, education, and personality traits predict attainment in adult occupational prestige over 17 years. Journal of Vocational Behavior. 2012;81: 218–226. doi:10.1016/j.jvb.2012.07.005

54. Henrich J, Heine SJ, Norenzayan A. The weirdest people in the world? Behavioral and Brain Sciences. 2010;33: 61–83. doi:10.1017/S0140525X0999152X

55. Hauser RM, Warren JR. Socioeconomic indexes for occupations: A review, update, and critique. Sociological Methodology. 1997;27: 177–298. doi:10.1111/1467-9531.271028

56. Haynes SN, Richard D, Kubany ES. Content validity in psychological assessment: A functional approach to concepts and methods. Psychological Assessment. 1995;7: 238. doi:10.1037/1040-3590.7.3.238

57. Bernard HR. Research methods in anthropology: Qualitative and quantitative approaches. 5th ed. Rowman Altamira; 2011.

58. Quinlan M. Considerations for collecting freelists in the field: Examples from ethobotany. Field Methods. 2005;17: 219–234. doi:10.1177/1525822X05277460

59. Weller SC. Structured interviewing and questionnaire construction. In: Bernard HR, Gravlee CC, editors. Handbook of methods in cultural anthropology. 2nd ed. Rowman & Littlefield; 2015. pp. 343–390.

60. Smith JJ, Borgatti SP. Salience counts-and so does accuracy: Correcting and updating a measure for free-list-item salience. Journal of Linguistic Anthropology. 1997;7: 208–209. doi:10.1525/jlin.1997.7.2.208

61. Mulac A. Evaluation of the speech dialect attitudinal scale. Speech Monographs. 1975;42: 184–189. doi:10.1080/03637757509375893

62. Mulac A. Assessment and application of the revised speech dialect attitudinal scale. Communication Monographs. 1976;43: 238. doi:10.1080/03637757609375935

63. Zahn CJ, Hopper R. Measuring language attitudes: The Speech Evaluation Instrument. Journal of Language and Social Psychology. 1985;4: 113–123. doi:10.1177/0261927X8500400203

64. Gundersen DF, Perrill NK. Extending the “Speech Evaluation Instrument” to public speaking settings. Journal of Language and Social Psychology. 1989;8: 59–61. doi:10.1177/0261927X8900800105

65. Wells JC. Accents of English. Cambridge University Press; 1982.

66. Labov W, Ash S, Boberg C. The atlas of North American English: Phonetics, phonology and sound change. Walter de Gruyter; 2005.

67. Shackleton RG. Phonetic variation in the traditional English dialects: A computational analysis. Journal of English Linguistics. 2007;35: 30–102. doi:10.1177/0075424206297857

68. Milroy L. Britain and the United States: Two nations divided by the same language (and different language ideologies). Journal of Linguistic Anthropology. 2000;10: 56–89. doi:10.1525/jlin.2000.10.1.56

69. Burton ML, Nerlove SB. Balanced designs for triads tests: Two examples from English. Social Science Research. 1976;5: 247–267. doi:10.1016/0049-089X(76)90002-8

70. Litman L, Robinson J, Abberbock T. TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. Behavior Research Methods. 2017;49: 433–442. doi:10.3758/s13428-016-0727-z

71. Lubke GH, Muthén BO. Applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group

comparisons. Structural Equation Modeling. 2004;11: 514–534. doi:10.1207/s15328007sem1104_2

72. Mardia KV. Measures of multivariate skewness and kurtosis with applications. Biometrika. 1970;57: 519–530. doi:10.1093/biomet/57.3.519

73. Mardia KV. Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. Sankhyā: The Indian Journal of Statistics, Series B. 1974;36: 115–128.

74. Henze N, Zirkler B. A class of invariant consistent tests for multivariate normality. Communications in Statistics - Theory and Methods. 1990;19: 3595–3617. doi:10.1080/03610929008830400

75. Royston JP. An extension of Shapiro and Wilk's W test for normality to large samples. Applied Statistics. 1982;31: 115–124. doi:10.2307/2347973

76. Royston JP. Some techniques for assessing multivarate normality based on the Shapiro-Wilk W. Applied Statistics. 1983;32: 121–133. doi:10.2307/2347291

77. Bulmer MG. Principles of statistics. Courier Corporation; 1979.

78. West SG, Finch JF, Curran PJ. Structural equation models with nonnormal variables: Problems and remedies. In: Hoyle RH, editor. Structural equation modeling: Concepts, issues, and applications. Thousand Oaks, CA: Sage Publications; 1995. pp. 56–75.

79. Curran PJ, West SG, Finch JF. The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. Psychological Methods. 1996;1: 16–29. doi:10.1037/1082-989X.1.1.16

80. Kaiser HF, Rice J. Little jiffy, mark IV. Educational and Psychological Measurement. 1974;34: 111–117. doi:10.1177/001316447403400115

81. Nunnally JC. Psychometric theory. 2nd ed. New York: McGraw-Hill; 1978.

82. Tabachnick BG, Fidell LS, Osterlind SJ. Using multivariate statistics. Allyn & Bacon/Pearson Education; 2001.

83. Bartlett MS. Properties of sufficiency and statistical tests. Proceedings of the Royal Society A. 1937;160: 268–282. doi:10.1098/rspa.1937.0109

84. Holgado-Tello FP, Chacón-Moscoso S, Barbero-García I, Vila-Abad E. Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. Quality & Quantity. 2010;44: 153–166. doi:10.1007/s11135-008-9190-y

85. Savalei V. What to do about zero frequency cells when estimating polychoric correlations. Structural Equation Modeling. 2011;18: 253–273. doi:10.1080/10705511.2011.557339

86. Jöreskog KG, Sörbom D. PRELIS 2 user's reference guide: A program for multivariate data screening and data summarization: A preprocessor for LISREL. Scientific Software International; 1996.

87. Rhemtulla M, Brosseau-Liard PÉ, Savalei V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. Psychological Methods. 2012;17: 354. doi:10.1037/a0029315

88. DiStefano C, Morgan GB. A comparison of diagonal weighted least squares robust estimation techniques for ordinal data. Structural Equation Modeling. 2014;21: 425–438. doi:10.1080/10705511.2014.915373

89. Kline P. Intelligence: The psychometric view. Florence, KY: Taylor & Frances/Routledge; 1991.

90. Costello AB, Osborne JW. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical Assessment, Research & Evaluation. 2005;10: 1–9.

91. Ruscio J, Roche B. Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. Psychological Assessment. 2012;24: 282. doi:10.1037/a0025697

92. Hershberger SL. Tetrachoric correlation. Encyclopedia of Statistics in Behavioral Science. John Wiley & Sons, Ltd; 2005. doi:10.1002/0470013192.bsa676

93. Kaufman L, Rousseeuw PJ. Partitioning around medoids (Program PAM). Finding groups in data: An introduction to cluster analysis. New York: John Wiley & Sons; 1990. pp. 68–125. doi:10.1002/9780470316801.ch2

94. Duda RO, Hart PE, Stork DG. Pattern classification and scene analysis. New York: Wiley; 1973.

95. Dunn JC. Well-separated clusters and optimal fuzzy partitions. Journal of Cybernetics. 1974;4: 95–104. doi:10.1080/01969727408546059

96. Linton R. The study of man. New York: Appleton-Century-Crofts; 1936.

97. Schriesheim CA, Hill KD. Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. Educational and Psychological Measurement. 1981;41: 1101–1114. doi:10.1177/001316448104100420

98. Harrison DA, McLaughlin ME. Cognitive processes in self-report responses: Tests of item context effects in work attitude measures. Journal of Applied Psychology. 1993;78: 129–140. doi:10.1037/0021-9010.78.1.129

99. Hothorn T, Hornik K, Van De Wiel MA, Zeileis A. A Lego system for conditional inference. The American Statistician. 2006;60: 257–263. doi:10.1198/000313006X118430

100. Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. Developmental Review. 2016;41: 71–90. doi:10.1016/j.dr.2016.06.004

101. Jorgensen TD, Kite BA, Chen P-Y, Short SD. Finally! A valid test of configural invariance using permutation in multigroup CFA. The Annual Meeting of the Psychometric Society. Springer; 2016. pp. 93–103. doi:10.1007/978-3-319-56294-0_9

102. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural Equation Modeling. 2007;14: 464–504. doi:10.1080/10705510701301834

103. Lumley T. Complex surveys: A guide to analysis using R. John Wiley & Sons; 2011.

104. Oberski DL. lavaan.survey: An R package for complex survey analysis of structural equation models. Journal of Statistical Software. 2014;57: 1–27. doi:10.18637/jss.v057.i01

105. United States Census Bureau, American FactFinder. B02001: Race. In: 2012-2016 American Community Survey [Internet]. 2016 [cited 7 Dec 2017]. Available: https://factfinder.census.gov/

106. United States Census Bureau, American FactFinder. P2: Urban and rural. In: 2010 US Census [Internet]. 2010 [cited 20 Sep 2017]. Available: https://factfinder.census.gov/

107. United States Census Bureau, American FactFinder. S1501: Educational attainment. In: 2012-2016 American Community Survey [Internet]. 2016 [cited 7 Dec 2017]. Available: https://factfinder.census.gov/

108. Office for National Statistics, Nomis. KS201UK: Ethnic group. In: 2011 UK Census [Internet]. 2011 [cited 20 Sep 2017]. Available: https://www.nomisweb.co.uk/

109. Department for Environment, Food & Rural Affairs. Rural population 2014/15. In: UK Rural Population and Migration Statistics [Internet]. 2012 [cited 20 Sep 2017]. Available: https://www.gov.uk/government/publications/rural-population-and-migration/rural-population-201415

110. Office for National Statistics, Nomis. QS501UK: Highest level of qualification. In: 2011 UK Census [Internet]. 2011 [cited 7 Dec 2017]. Available: https://www.nomisweb.co.uk/

111. Satorra A, Bentler PM. Corrections to test statistics and standard errors in covariance structure analysis. In: von Eye A, Clogg CC, editors. Latent variables analysis: Applications to developmental research. Thousand Oaks, CA: SAGE Publications, Inc.; 1994. pp. 399–419.

112. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling. 1999;6: 1–55. doi:10.1080/10705519909540118

113. Yu C-Y. Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. University of California, Los Angeles. 2002.

114. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychological Bulletin. 1955;52: 281. doi:10.1037/h0040957

115. Lawshe CH. A quantitative approach to content validity. Personnel Psychology. 1975;28: 563–575. doi:10.1111/j.1744-6570.1975.tb01393.x

116. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. New Jersey: Lawrence Erlbaum Associates; 1988.

117. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research. 1981;18: 39–50. doi:10.2307/3151312

118. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science. 2015;43: 115–135. doi:10.1007/s11747-014-0403-8

119. Voorhees CM, Brady MK, Calantone R, Ramirez E. Discriminant validity testing in marketing: An analysis, causes for concern, and proposed remedies. Journal of the Academy of Marketing Science. 2016;44: 119–134. doi:10.1007/s11747-015-0455-4

120. Skrondal A, Rabe-Hesketh S. Prediction in multilevel generalized linear models. Journal of the Royal Statistical Society: Series A. 2009;172: 659–687. doi:10.1111/j.1467-985X.2009.00587.x

121. Krippendorff K. Content analysis: An introduction to its methodology. 3rd ed. Thousand Oaks, CA: Sage; 2012.

122. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychological Methods. 1996;1: 30–46. doi:10.1037/1082-989X.1.1.30

123. Fleiss JL, Fleiss JL. Reliability of measurement. Design and analysis of clinical experiments. John Wiley & Sons; 1986. pp. 1–32.

124. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16: 297–334. doi:10.1007/BF02310555

125. Revelle W, Zinbarg RE. Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. Psychometrika. 2009;74: 145–154. doi:10.1007/s11336-008-9102-z

126. Lance CE, Butts MM, Michels LC. The sources of four commonly reported cutoff criteria: What did they really say? Organizational Research Methods. 2006;9: 202–220. doi:10.1177/1094428105284919

127. Tol P. Colour Schemes [Internet]. Utrecht, Netherlands: SRON; 2018 May. Report No.: SRON/EPS/TN/09-002. Available: https://personal.sron.nl/~pault/data/colourschemes.pdf