Adjusting for Sample Bias in VAA Datasets: Can it be done and does it matter?

Jonathan Wheatley and Micha Germann

Abstract

Data generated from VAAs on the political orientations of users have increasingly been used in place of survey data to explain political phenomena such as voter behaviour and the ideological positioning of party supporters. However, the use of VAA-generated data in this way is controversial due to the issue of sample bias. This paper explores the relative merits of three different methods to deal with this bias: raking, propensity score adjustment and entropy balancing. It applies these methods to a dataset generated from the EUVox VAA in Spain deployed in the run-up to the 2014 elections for the European Parliament. Data from the 2009 European Elections Survey are used as a reference sample. I find that none of the three methods can adequately compensate for sample bias, probably because of the rather small number of common items used to balance the two samples. However, by drawing on another instance of the EUVox VAA, this time from England, the paper goes on to argue that meaningful research using VAA-generated data can be carried out, even if this data is not fully representative.

Introduction

The primary aim of Voting Applications (or VAAs) is to help voters decide which party or candidate most closely matches their policy preferences; a significant body of literature therefore relates to issues involving VAA design, such as the spatial models and metrics used in them (Louwerse and Rosema 2011; Germann et al. 2014), the effects of statement selection (Walgrave et al. 2009) or methods for coding parties and candidates (Trechsel and Mair 2011). A second—and equally important—strand of research focusses on the effects of VAAs on politics, specifically their potential impact on voting behaviour (Fivaz, Pianzola and Ladner 2010), the extent to which they may affect voter turnout (Marschall and Schultze 2012; Germann and Geminis 2014) or the possibility that populist parties may exploit VAAs by adapting to the policy positions of the average voter (Ramonaitė 2010).

A third strand of the research goes beyond the issue of VAA design and draws from the data provided by VAA users. VAAs can attract many users and thereby generate very large datasets containing their policy preferences. Most VAAs also include supplementary questions in which users are asked to provide demographic data such as age, gender and education, as well as data relating to their political orientation. This can then be accessed by the researcher. It is hoped that VAA-generated data can be used to investigate some of the central questions of political science, on topics such as voting behaviour and party affiliation. One example of how VAA-generated data can be exploited in this way is in identifying latent ideological dimensions and to map party supporters on an ideological space (Wheatley et al 2012; Wheatley 2012; Mendez and Wheatley 2014). In short, VAAs can play the role of a kind of public opinion survey, which is significantly less costly than face-to-face surveys.

However, in order to use the VAA as a survey, we need to be sure that data it generates is more or less representative of the population at large. This condition does not appear to hold. Overall, VAA users tend to be more well-educated, younger and more left-wing than the rest of the population (Wheatley et al 2012; Wheatley 2012; Mendez and Wheatley 2014). Typically also around two-thirds of them are male. To what extent does this lack of representativeness preclude the possibility that VAA-generated data can be exploited for meaningful research?

Of course, even telephone surveys and face-to-face surveys such as opinion polls are not always based on a fully representative sample. For this reason, raw opinion data obtained in the poll is often weighted to compensate for sampling bias. The principle behind weighting is that individuals who belong to a category (typically, but not always, a demographic category) that is over-represented receive a lower weight than those who belong to an under-represented category.

If weighting could be used to make VAA-generated data more or less representative of the general population it could provide rich benefits not only for academics—who would now have a readily available “survey tool”—but also for VAA practitioners. If VAAs could act as a kind of proxy opinion poll that could provide insights into the voting intention of the population at large, they are likely to generate significant media interest, which would in turn generate strong public interest in the tool.

This paper explores the effectiveness of three forms of weighting with the aim of making VAA-generated data more representative: raking, propensity score analysis (PSA) and entropy balancing. It uses as its sample data from the EUvox VAA that was obtained from Spain in the run-up to the 2014 European elections and included more than 100,000 respondents. For each form of weighting, benchmark covariate distributions for the VAA-generated sample are drawn from a (more-or-less) representative reference sample made up of Spanish respondents from the 2009 European Elections Survey (EES). The weighting is designed to compensate for biases in education, gender, age, political interest and self-placement on a left-right ideological scale and the effectiveness of the weighting process is evaluated according to whether it produces distributions of two test variables—vote intention in the European elections and an attitudinal variable on the need for immigrants to adapt to the culture of the host nation—that correspond more or less to the overall voting population (in the first case) and to the reference sample (in the second).

The paper proceeds as follows. First, I briefly explain both Spanish EUVox and EES datasets in terms of how they were generated and how representative they were. I then outline the three weighting procedures before applying them to the datasets and evaluating their effectiveness. Finding them to be by and large ineffective in producing a distribution on the test variables that corresponds to the distributions amongst the voting population and the reference sample, I then ask whether this apparently insurmountable problem precludes us from carrying out meaningful research on VAA-generated data. By comparing data analysis on dimensionality on two very different subsets of a second dataset (the EUVox dataset from the UK), I conclude on a cautiously optimistic note: that while VAA datasets are not representative, this does not always matter.

The Euvox Voting Advice Application

The EUvox VAA was deployed in twenty-seven out of twenty-eight EU countries (all except Malta) in the run-up to the 2014 elections to the European Parliament, generating around a million users. Respondents who filled in the VAA were asked to give their opinion on a set of thirty policy statements that were designed to cover a wide range of topics relating to economic, cultural and EU-related issues. For each policy statement respondents were offered a five-point response scale: “Completely Agree”, “Agree”, “Neither Agree nor Disagree”, “Disagree” and “Completely Disagree”. In addition a user could indicate “No Opinion”. Respondents were also asked to complete a number of supplementary questions that included gender, age, education, vote intention and interest in politics and to place themselves on a left-right scale. Although these questions were voluntary, most users completed them.

The most widely used of the EUVox websites was the Spanish site, which received 289,552 completed entries from users. After extensive cleaning 149,536 unique users were identified.[1] However, these were far from representative of the Spanish population. Of 106,951 users who answered the question on how they intended to vote in the forthcoming European elections, 32,088 professed themselves to be undecided and 7,363 said they would not vote. Of the remaining 67,500 who expressed a valid vote

1 For cleaning I removed: 1) all cases in which the time taken to complete the 30 issue statements of the VAA was less than 120 seconds; 2) all cases in which the time to respond to any one issue statement was less than two seconds; 3) all cases in which the time taken to respond to three or more issues statements was less than 3 seconds; 4) all cases in which the respondent answered ten successive issue statements in the same way; 5) all cases in which the user completed the questionnaire by smart phone (this is because it may not have been intuitively obvious about how to resister "no opinion" by smart phone). I then (6) sorted the data by an anonymised code that corresponded to the IP address (first), date of birth (second), gender (third), before removing all consecutive items with the same IP address unless it was obvious from date of birth or gender that they were different users (if the user did not specify date of birth and gender they would be deleted). Finally I (7) removed all those who self-identified with a citizenship other than UK citizenship and all those that claimed a date of birth prior to 1920 (on the grounds that they were probably fictitious entries).

intention, their voting intention in percentage terms is provided in Table 1, together with the actual percentages of votes obtained by each party in the election itself.

From Table 1 we can see that EUVox users were far more likely to vote for the left-wing blocs La Izquierda Plural, Podemos and Primavera Europea than the average voter. Certain demographic categories were also over-represented (or under-represented) in the Euvox sample; thus 66.31% of the sample were male, their average age was 34 and 56.81% of those declaring their educational attainment had university education. EUVox users were therefore predominantly more male, more well-educated and younger than the overall population.

[Insert Table 1 about here]

Five of the supplementary questions (on gender, age, education, vote intention, interest in politics and left-right self-placement) were scaled in precisely the same way as corresponding items in the European Election Survey (see below), as responses to these questions were used as benchmark covariates.

The European Elections Survey

The reference sample that I use to provide the benchmark covariate distributions for VAA-generated sample is the dataset obtained from the 2009 European Election Survey (EES) in Spain, carried out immediately after the 2009 European elections, which consisted of the responses of 1,000 Spanish citizens. The sample of EES respondents was designed in such a way as to be more or less representative of the population at large. Of the 1,000 adult respondents, 453 were male and 547 female, the average age was 46 and 363 declared they had received higher education. In terms of party support, of those who said they had cast a valid vote for one of the parties standing in the European elections, 45.3% said they had cast their vote for the Partido Popular (PP) and 42.0% for Partido Socialista Obrero Español (PSOE). This compares to actual voting figures of 42.1% and 38.8% respectively. Both in terms of demographic group and in terms of vote intention, the EES survey sample seems to be more or less representative of the Spanish population.

However, because the EES was not fully representative, it also assigned each respondent a weight that was based on gender, age, region of residence, turnout in both the 2009 European elections and in the (2008) national election. As with all forms of survey weighting, respondents belonging to groups that were under-represented in the survey were assigned proportionally larger weights than those belonging to over-represented groups.

Of course, ideally the reference sample should be "gathered reasonably contemporaneously with the non-random sample data" (Watson and Elliot 2012). A five-year gap may be stretching the limits in terms of what constitutes "reasonably contemporaneously". The relative strengths of certain categories, particularly categories such as education level, may have varied somewhat between 2009 and 2014. Nevertheless, we do not expect these changes to be to great enough to invalidate the comparison; we still expect the relevant distributions of the benchmark covariates (gender, age, education, political interest and left-right self-placement) observed in the 2009 EES dataset to approximate the distributions observed across the population as a whole in 2014.

In addition to questions involving the benchmark covariates, respondents to the EES also answered a series of questions about their political views. One such question was later virtually replicated in the EUVox VAA. This was "Immigrants should adapt to the customs of Spain" (appearing in the form "Immigrants should adapt to the values and culture of Spain" in EUVox).

Post-Stratification Weighting: Three Methods

In this paper, I use three methods of post-stratification weighting: raking, propensity score analysis (PSA) and entropy balancing. The principles behind these three methods are the following:

Raking is an iterative procedure that compares the marginal distributions of the benchmark covariates (in our case age, gender, education, political interest and left-right self-placement) in a representative dataset (in our case the EES survey or reference sample) with the distributions of the same covariates in a non-randomly sampled dataset (e.g. VAA data) and generates weights for each unique combination of covariates according to whether—and the extent to which—that particular combination is over-represented or under-represented in the latter dataset with respect to the former. The result is that every respondent in the non-randomly sampled dataset is assigned a weight corresponding to his or her scores on the benchmark covariates, and specifically according to how frequently that combination crops up in the reference sample.[2] While typically raking, like other post-stratification weighting techniques, relies on census-based known population totals to provide the expected distributions of the required covariates, in this paper we use the distributions from the reference sample in order to compare the effectiveness of raking weighting techniques with those of PSA and entropy balancing.

One problem with raking is the so-called "dimensionality curse" which increases in proportion to the number of benchmark covariates (or dimensions). Given a large number of covariates, certain permutations of covariates may crop up very rarely in the reference sample, meaning that calculating a weight for that combination may be problematic. One solution is to substitute these many dimensions with a one-dimensional probability in what is known as *propensity score analysis* (PSA). PSA merges

2 For a very clear overview of the raking procedure, see Battaglia, Michael P., Izrael, David, Hoaglin, David C. and Frankel, Martin R. Tips and Tricks for Raking Survey Data (a.k.a. Sample Balancing) (American Association for Public Opinion Research 2004) at <http://www.amstat.org/sections/srms/Proceedings/y2004/files/Jsm2004-000074.pdf>, accessed 24 July 2014.

the reference sample (known as the “treatment group”) with the non-random sample (the “control group”) into a single dataset, creates a dummy variable which is set to 1 for the treatment group and to 0 for the control group and performs a logit or probit regression with the dummy variable as the dependent variable and the benchmark covariates as the independent variables. It thereby generates a set of *propensity scores* for all observations in both datasets, corresponding to the probability that the observation in question will belong to the treatment group. The propensity scores can then be converted into weights. Lee, Lessler and Stuart do this by assigning all entries in the treatment group a weight of 1 and all those in the control group a weight of p/(1-p), where p is the propensity score (Lee, Lessler and Stuart 2011). In this paper I use the same method.

In its usual form PSA imposes constraints that will lead to a convergence of the first moments of the covariate distributions, i.e. the means of the benchmark covariates in the treatment group and the reweighted control group should converge. However, these constraints are unlikely to lead to convergence of the second and third moments, i.e. after weighting, the variances and skewnesses will probably not converge. To adjust for this it is possible to also include in the regression analysis the squares and cubes of all covariates except for the dichotomous variables. We also include the first-order interaction terms (obtained by multiplying each covariate by each of the others in turn) to adjust for the joint distributions of these variables.

The third method employed in this paper is *entropy balancing*. Entropy balancing uses a reweighting scheme to generate a set of weights for the control group so that the treatment group and reweighted control group satisfy a large set of balance constraints (Watson and Elliot 2013). These balance constraints ensure that the distributions of covariates in the two groups exactly match on all pre-specified covariate moments, typically including not only the first moment (the mean), but also the second moment (the variance) and often also the third moment (skewness) (Hainmueller 2012).

Entropy balancing software can also allow the researcher to introduce weights for the treatment group, an undertaking that is problematic for PSA and raking.

Analysing the Datasets

The three above-mentioned procedures were performed on user data from the Spanish EUVox VAA (the control group) and from the 2009 European Elections Survey (the treatment group or reference sample). For all three methods weights were generated by controlling for distributions on five benchmark covariates: gender, age, education, political interest and left-right self-placement. For PSA, we perform two analyses: the first using a simple logit regression to generate propensity scores from the five benchmark covariates, and the second including the squares and cubes of all non-dichotomous variables in the regression, in order to factor in the higher order moments (variance and skewness), as well as the first-order interaction terms. For entropy balancing, we also perform two analyses: first using the treatment group (i.e. the EES dataset) without weights and second incorporating the EES's own weights. Both adjust for the means of the covariates (first moment), variances (second moment) and skewness (third moment). Raking was performed using the *rake* function in the “survey” package in R; PSA used the *matchit* function in the “Matchit” package in R; while for entropy balancing the “ebalance” package in STATA was used.

The effect of these procedures on the sample characteristics of the benchmark covariates after each weighting procedure is used is shown in Tables 2 a-f, which compare the means and standard deviations of the benchmark covariates of the reference sample with those of the reweighted control group (i.e. the EUVox sample). In all cases the means converge fairly well and in four cases (the case in which raking is used, the more complex version of PSA and the two cases in which entropy balancing is used), the standard deviations also converge rather well. In the cases in which raking and entropy balancing are used, the convergence appears to be more or less perfect.

[Insert Table 2 about here]

The left hand side of Figure 1 shows a plot of the standardized bias for all benchmark covariates (including squares, cubes and interaction effects) in the unadjusted EUvox data and after each of the reweighting methods. This measures the difference in means between the treatment (reference) and control group, scaled by the standard deviation of the former. The right hand side of Figure 1 shows the p values for two-sided t-tests for the differences in means of all benchmark covariates using the (EUVox) data both unweighted and after the various reweighting methods have been applied. The dotted line near the left of the diagram represents the 0.1 level of significance; all icons to the left of the line indicate a difference in means that is significant at least at the $p<0.1$ level. Figure 1 suggests that only the standard form of PSA (i.e. not taking into account the higher moments) fails to bring about sufficient convergence, and only then with respect to the second and third moments and interaction terms of certain covariates. In no other cases do the means and the higher moments (standard deviation and skewness) of the reweighted samples differ to a statistically significant degree from those of the reference sample. We can thus say that virtually all of our reweighting methods successfully balance the characteristics of the benchmark covariates with those of the reference sample, although Figure 1 suggests that entropy balancing is clearly superior.

[Insert Figure 1 about here]

However, the real test is how the distributions of the two test covariates that are **not** benchmark covariates compare after weighting. These two variables are vote intention in the European parliamentary elections and concordance with the statement “immigrants should adapt to the values and culture of Spain”. We hypothesise that if the reweighting process were to simulate a truly representative sample, then the vote intention variable after reweighting should distributed in a similar way to how Spanish citizens actually voted. Similarly, we would expect that the distribution of opinion on the attitudinal variable on immigration, once reweighted, should approximate the distribution observed for the same variable in the EES, notwithstanding the possibility that attitudes towards immigration may have changed somewhat since 2009.

Table 3 shows the distributions of the vote intention variable on the EUVox sample both unweighted and after all reweighting procedures have been applied, together with the way the Spanish electorate actually voted in the European elections of 25 May 2014.[3] In each case the expected frequency of voting for a particular party after reweighing is followed by two numbers in brackets which represent the lower and upper 95% confidence bounds. From Table 3, we can see that there are significant discrepancies between the frequencies observed after reweighting and the way the electorate voted. In particular, like the unweighted sample, the proportions assigned to the two main parties, the Partido Popular (PP) and the Partido Socialista Obrero Español (PSOE), in the reweighted samples significantly under-estimate their true electoral strength. At the same time, the electoral strength of the minor parties, both Left and Right, but particularly on the Left, are exaggerated to a significant extent by the reweighted EUVox samples, just as they are by the unweighted EUVox sample. The parties most affected by this exaggeration are the left-wing parties Podemos and Primavera Europea, and to a lesser extent the left-wing Izquierda Plural, the centrist Ciudadans and the right-wing VOX. The explanation for this cannot be that support for these parties dipped significantly in the last days of the election campaign, giving the main parties an unexpected boost at the last minute; on the contrary, the big surprise was the unexpected surge of Podemos on election day. In the run-up to the elections, opinion polls gave Podemos a share of the vote of between 1.0 and 3.5%, compared with their final tally of 8%.

3 The discrepancies observed between the second column of Table 2 and Table 1 are due to the fact that the frequencies given in Table 1include *all* those who gave information about their vote intention (N = 106,951), while Table 2 includes only those who provided information on *all* benchmark covariates *and* who stated their intention to vote for a particular party (i.e. were not undecided).

[Inset Table 3 about here]

Similar discrepancies are revealed if we compare the distribution of responses for the attitudinal variable on immigration in the EES sample (the treatment group) with those in the EUVox sample after different reweighting procedures have been applied. Table 4 shows the distributions observed in the EES sample both with and without weighting, as well as the distributions. We can see that whatever weighting procedures are used, the distributions observed for the reweighted EUVox datasets do not get sufficiently close to the treatment group and significant differences remain. We could argue that the difference relates to the five years that passed between 2009 and 2014 and that Spanish citizens have become more accommodating in their attitudes towards immigrants. However, this hypothesis is counter-intuitive, given the depth of economic crisis that engulfed Spain in the intervening years and the tendency for anti-immigrant sentiment to run high in times of economic crisis.

[Inset Table 4 about here]

Overall, we can only conclude that *no* post-stratification procedure has been able to replicate adequately the covariate distributions observed in the EES survey or, indeed, those observed in the population at large. In part, this may be due to the fact that we have used as a reference sample a survey that was conducted five years before the VAA. However, it is unlikely that the distributions of the benchmark covariates have changed sufficiently drastically over the five-year period to produce the sort of discrepancies that we observe. More likely, the reason for the divergence relates to the relatively small number of benchmark covariates that are used to balance the treatment and control groups. Ideally, for all post-stratification weighting procedures as large a number of benchmark covariates as possible are used. In our case, this should include covariates such as income, region of residence, settlement type, former voting behaviour and political knowledge. In terms of VAA design, this would imply that if we want to use VAA-generated data to draw conclusions about the population at large, we would need to include a larger repertoire of supplementary questions that exactly replicate those incorporated into a current sociological survey from which data is readily available. This is an ambitious undertaking, both because VAA users may find answering many supplementary questions wearisome and intrusive and because often the relevant survey data is simply not available.

Does it matter?

Because it is highly problematic to reconstruct a representative sample of the population from non-random sample data generated from a VAA, does this mean that VAA-generated data can tell us nothing? The previous section would suggest that using VAA-generated data as a kind of surrogate opinion poll simply will not work, at least not unless we have at our disposal a battery of data from a long list of critical supplementary questions as well as corresponding representative survey data. Should we not then discard VAA-generated data as somehow “contaminated”?

My own view is that whether VAA-generated data is useful or not depends on the question the researcher is trying to ask. Obviously, if the researcher is trying to find out the voting intention of the electorate, the VAA is not the best tool to find this out. I will end this paper with an example of how VAA data may still be useful in addressing one of the central questions of political science; how to identify the ideological dimensions that define the political space.

Typically scholars who have investigated ideological dimensions have focused their attention on the official position of parties and party elites. Thus, the Manifesto Research Group/Comparative Manifestos Project (MRG/CMP) have been collecting manifesto data since 1979 and and claim to have carried out content analyses of party election programmes from more than 50 countries going back to

1945 (Budge et al., 1987, 2001; Klingemann et al., 2006). Using this data, Budge, Robertson and Hearl use a two stage factor analysis to identify latent ideological dimensions from data generated by this method (Budge et al., 1987). Another approach is to invite independent experts to code parties according to a set of already defined policy dimensions (Benoit and Laver, 2006; Castles and Mair, 1984; Huber and Inglehart, 1995; Marks et al., 2006). More recently the Chapel Hill team (based at the University of North Carolina) has adapted this approach by subsequently performing confirmatory factor analysis (CFA) on the pre-defined dimensions using data drawn from the Chapel Hill Expert Survey of European parties (Bakker, Jolly and Polk 2012). A third approach is to use elite surveys in which politicians and party functionaries are asked to position their own parties in terms of policy (Kitschelt et al., 1999). It is relatively rare that public opinion survey data is used to investigate dimensionality; an exception to this is Sani and Sartori (1983), who drew data from a mass survey in order to measure party polarization based on the self-defined location of party supporters.

VAA data is ideal for testing theories of dimensionality, because typically VAAs allow respondents to record their views on a large number of issues (more than is often feasible in face-to-face or telephone surveys), and latent dimensions can be identified inductively by means of factor analysis and other dimension reduction techniques. Such analysis has already been applied to a number of VAA-generated datasets and have suggested that both the number and nature of latent ideological dimensions (in terms of which issue "belongs" to each dimension) may be context-dependent (Wheatley et al 2012; Wheatley 2012; Mendez and Wheatley 2014). The argument I intend to make here is that this approach is promising, both because research into dimensionality does not require a representative survey (given that elite surveys are frequently used) and because dimension reduction techniques seem to be remarkably robust and produce similar results when applied to highly dissimilar groups of respondents.

The example I use here is the result of analysis of data generated by the EUVox VAA in England that was also operational in the run-up to the May 2014 European parliamentary elections. In total data retrieved from the English Euvox site contained 131,040 completed entries from users of which 70,201 remained after extensive cleaning. The users of the England VAA differed from those of many other VAAs in terms of the demographic and ideological profile of its users. While overall VAA users tend to be young and left-leaning (as was the case in the Spanish version of EUVox, see above), in England an i-frame of the VAA was incorporated into the website of the popular newspaper the Daily Mail, which tends to have an older, more right-wing readership. An i-frame was also placed in The Guardian's website, which attracts younger, more left-wing readers.

The characteristics of 1) the overall group of VAA users together with 2) those who accessed the site via the Daily Mail website and 3) those who accessed the site via the Guardian website are shown in Table 5. We can see that the Daily Mail sample was considerably older (median date of birth, 1969) than the Guardian sample (1980). The Guardian sample were also slightly more interested in politics than the Daily Mail sample. Moreover, the Guardian sample were, on average, considerably more well-educated than the Daily Mail sample; of those who declared their educational level, just over 77% of the Guardian sample claimed higher education compared with just over 49% of the Daily Mail sample. The fact that 37% of the latter sample did not declare their educational level at all (compared with 12% of the Guardian sample) suggests that the real figure for this group may be lower. Even more importantly, Daily Mail readers were far to the right of Guardian readers on a scale from Left (0) to Right (10) and more conservative on a scale from socially liberal (0) to socially conservative (10). On average, the Daily Mail sample scored 6.2 on the first count and 6.1 on the second, while the Guardian sample averaged out at 3.8 and 2.8 respectively. In terms of vote intention, a massive 61% of the Daily Mail sample, but just 13% of the Guardian sample pledged to vote for the United Kingdom Independence Party (UKIP). In all, over 80% of the Daily Mail sample pledged to vote for a right-wing party (the Conservatives or UKIP), while just 23% of the Guardian sample pledged to do so.

In order to test the dimensionality of the political space in England as defined by VAA users, I first perform a factor analysis on *all* VAA users to identify latent dimensions. I also use Mokken Scale Analysis to test whether the dimensions identified form valid scales that conform to the monotone homogeneity model (Mokken, 1971; Sijtsma and Molenaar, 2002; van der Ark et al., 2008). I then use confirmatory factor analysis (CFA) and a repeat round of Mokken Scale Analysis to test whether the dimensional structure identified in the first analysis adequately fit the two sub-samples.

The method I use to apply factor analysis is that suggested by Gerbing and Anderson (1988), which entails first using exploratory factor analysis (EFA) as a preliminary technique to identify latent dimensions[4], then to use CFA to evaluate and, if necessary refine, the scales identified by EFA. EFA provides a first insight into how the latent dimensions (if such exist) are structured. I then group together items that load strongly (with factor loadings greater than 0.4) or unambiguously (with factor loadings greater than 0.3) onto one or other dimension and perform CFA in order to test the hypothesis that the items in each group are indeed associated with a specific latent factor.[5] As long as the Comparative Fit Index (CFI) remains below 0.95 and the root mean square error of approximation (RMSEA) remains above 0.08, I examine the modification indices in order to identify any item that appears to load strongly onto a factor associated with another group or exhibits abnormal error covariance with other items, remove that item and retest the model with CFA. I repeat this process until the RMSEA and the CFI are within the specified limits.[6] Finally, I perform Mokken Scale Analysis to

4 A “principal axis” factor solution is applied with a promax rotation and the R package “psych” is used to perform the analysis. The function fa.poly first identifies polychoric correlations between categorical variables before performing the factor analysis.

5 To perform CFA, I use the package “lavaan” in R.

6 Researchers often disagree as to what constitutes a good or adequate fit in terms of the thresholds of these indices. The thresholds I will use for the purpose of this paper are mid-way between the most stringent (RMSEA<0.05, CFI>0.95, TFI>0.95) and the most permissive (RMSEA<0.1, CFI>0.90, TFI>0.90) criteria. Based on guidelines suggested by Hu and Bentler (1999), Browne and Cuddick (1993) and Brown (2006), I propose that there is an adequate fit if RMSEA is less than 0.08, and both the CFI and TFI are more than or near to 0.95.

test whether the groups identified form Mokken Scales and satisfy monotone homogeneity criteria.[7] A lower bound for the item scalability coefficients (H_i) was set at 0.3; items that either produced a value of H_i below 0.3 or did not satisfy the monotone homogeneity model were removed from the scale.

Factor analysis suggests that a three-dimensional structure is the best way to define the policy space of the full sample of VAA users. This was deduced by examining the scree plots, which show the eigenvalues of each component when EFA is applied, and identifying a break-point or “elbow” where the curve flattens out. The scree plot is shown in Figure 2. The issue statements that load onto each dimension are shown in Table 6 together with the item scalability coefficients (H_i) obtained when Mokken Scale Analysis is applied, as well as the overall test scalability coefficient H. At the bottom of the page the CFI, Tucker Lewis Index (TLI) and RMSEA are shown. All are within the specified limits (CFI>0.95, RMSEA<0.08). Sixteen of the thirty issue statements proved scalable: One scale appears to measure tolerance towards outsiders, with attitudes towards the EU, immigration and gay rights loading on to this dimension, another seems to be an economic left versus economic right scale, while the third includes just two items relating to freedom of information and freedom to protest.

Repeating the analysis first on the Daily Mail sample and then on the Guardian sample (see Tables 7 and 8), we see that an identical dimensional structure can be applied to these two samples without compromising the goodness-of-fit indices and without violating either the H_i>0.3 threshold or the monotone homogeneity criteria when Mokken Scale Analysis is applied. The only difference observed is that the scales are slightly weaker in the Daily Mail sample (lower values of H_i) than the in the overall sample.

This last analysis suggests that in order to identify latest dimensions from data generated by VAA users,

7 “Crit” values of more than 80 for any one item suggests that the item violates monotonicity requirements. To perform Mokken scale Analysis and to check monotonicity, I use the package “mokken” in R.

it may not be necessary for the data to be fully representative of the population as a whole. At least in the case of data from the EUVox VAA for England, the method seems extremely robust and produced very similar results for highly divergent sectors of the population. Of course, a critic could argue that despite their differences, all VAA users, both those who accessed the site via the Guardian and those who accessed it via the Daily Mail may have had a keener interest and knowledge of politics than the average citizen – and as such constitute “armchair experts”. Nevertheless, even if this were true, it would hardly make them an unsuitable panel for testing dimensionality, given that many studies of dimensionality indeed rely on experts.

Conclusion

Evidence from the Spanish data would suggest that, amongst all the methods of post-stratification weighting that were tested, entropy balancing achieves the best results, insofar as it produces a near-perfect convergence of mean, standard deviation and skweness amongst the benchmark varianbles. However, none of these methods was able to reproduce distributions on non-benchmark covariates that were similar to those observed either in a representative survey or amongst the population at large. It is likely that in order to compensate properly for sample selection bias, many more benchmark variables would be needed, whatever method for post-stratification weighing is used. Thus, the failure to achieve satisfactory results does not reflect upon shortcomings in the methods themselves; the method can only be as good as the data we feed it.

In practical terms, these obstacles may be difficult to circumnaviagate. The number of benchmark covariates available is conditioned by the need to avoid antagonising VAA users with a barrage of supplementary questions. Similarly, the frequent lack of availability of contemporary survey data may mean that we have to resort to using older and less reliable surveys to provide us with “representative” distributions. This means that efforts to use VAA-generated data to somehow mimic or replace survey data are likely to be in vain.

On the other hand, this does not mean that VAA-generated data cannot be exploited to investigate some of the core questions in political science. As the example of dimensionality shows, data does not always have to be representative to produce meaningful results. Certain findings can be quite robust when replicated on highly divergent samples. Furthermore, even highly unrepresentative VAA data can tell us something, if we treat VAA-users as a kind of “expert panel” and are explicit that our research does not require a representative sample.

In terms of the future direction of VAA development, the findings presented in this paper suggest that wherever possible, a representative survey should be carried out simultaneously with the launch of a VAA, with a significant number of key items common to both survey and VAA. Only then would it be possible to generalise our findings. However, this would require financial and human resources that are not always available.

Bibliography

Austin, Peter C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*, 46: 399–424.

Bakker, R. , Jolly, S. and Polk, J. (2012). Complexity in the European party space: Exploring dimensionality with Experts, *European Union Politics*, 13, pp. 219-245.

Benoit K and Laver M (2006). *Party Policy in Modern Democracies*. London: Routledge.

Brown, Timothy A. (2006). *Confirmatory Factor Analysis for Applied Research* (New York: Guildford Press).

Browne, M. W., and Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (eds), *Testing Structural Equation Models* (Newbury Park, CA: Sage), pp. 136-162.

Budge I, Robertson D and Hearl D.J. (eds) (1987). Ideology, Strategy, and Party Change: *Spatial Analyses of Post-war Election Programmes in 19 Democracies*. Cambridge: Cambridge University Press.

Budge I, Klingemann H.D., Volkens A, Bara J, Tanenbaum E, Fording RC, Hearl DJ, Kim H-M, McDonald M and Mendez S (2001). *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments 1945–1998*. Oxford: Oxford University Press.

Castles F.G. and Mair P. (1984). Left–right political scales: Some ‘expert’ judgements. *European*

Journal of Political Research, 12(1): 73–88.

Fivaz, J., Pianzola J. and Ladner A. (2010). More than Toys: A First Assessment of Voting Advice Applications' Impact on the Electoral Decision of Voters. *Working Paper No. 48, National Centre of Competence in Research (NCCR), Challenges to Democracy in the 21st Century*. Available online at: http://www.nccr-democracy.uzh.ch/publications/workingpaper/pdf/WP_48.pdf (accessed on 11 May 2014).

Gerbing, David W. and Anderson, James C. (1988) An updated Paradigm for Scale Development Incorporating Unidimensionality and its Assessment. *Journal of Marketing Research,* XXV, pp. 186-192.

Germann, M., Mendez, F., Wheatley, J. and Serdült, U. (2014). Spatial maps in Voting Advice Applications: The case for dynamic scale validation. *Acta Politica*, doi: 10.1057/ap.2014.3.

Germann, M. and Geminis K. (2014). Online gimmick or participation promoting tool? Smartvote and its Effect on Electoral Turnout, paper presented at the Annual Congress of the Swiss Political Science Association in Berne, January 30-31 2014.

Hainmueller, Jens (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis,* 20(1): 25-46.

Hu, Li-tze and Bentler, Peter M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), pp. 1-55.

Huber, J. and Inglehart R. (1995). Expert Interpretations of Party Space and Party Locations in 42 Societies. *Party Politics*, Vol. 1, No. 73, pp.73–111.

Kitschelt, H., Mansfeldova, Z., Markowski, R. and Toka, G. (1999) *Post-Communist Party Systems: Competition, Representation, and Inter-Party Cooperation.* Cambridge University Press, Cambridge.

Klingemann H.D., Volkens A., Bara J., Budge I. and McDonald M. (2006) *Mapping Policy Preferences. II: Estimates for Parties, Electors and Governments in Eastern Europe, European Union and OECD 1990–2003*. Oxford: Oxford University Press.

Lee, B., Lessler, J., and Stuart, E.A. (2011). Weight trimming and propensity score weighting. *PLoS ONE*, 6(3): e18174.

Marks, G., Hooghe, L., Nelson, M., and Edwards, E. (2006) Party Competition and European Integration in the East and West: Different Structure, Same Causality. *Comparative Political Studies*, 39(2), pp. 155-175.

Marschall, S. and Schultze, M. (2012). Voting Advice Applications and their Effect on Voter Turnout: The Case of the German Wahl-O-Mat. *International Journal of Electronic Governance*, 5(3), pp. 349-366.

Mendez, F. and Wheatley, J. (2014). Using VAA-generated Data for Mapping Party Supporters in the Ideological Space. In Garcia, D. and Marschall, S., *Matching Voters with Parties and Candidates: Voting Advice Applications in a Comparative Perspective*. Colchester: ECPR Press.

Mokken, R. J. (1971). *A Theory and Procedure of Scale Analysis*. Berlin, Germany: De Gruyter.

Sani G. and Sartori G. (1983) Polarization, fragmentation and competition in Western democracies. In: Hans D and Mair P (eds) *Western European Party Systems*. London: Sage, pp. 307–340.

Ramonaite, A. (2010). Voting Advice Applications in Lithuania: Promoting Programmatic Competition or Breeding Populism? *Policy and Internet*, 2(1), pp. 117–147.

Sijtsma, K. and Molenaar, I.W. (2002). *Introduction to Nonparametric Item Response Theory*. Thousand Oaks, CA: Sage.

Trechsel, A.H., and Mair P. (2011). When parties (also) position themselves: an introduction to the EU profiler. *Journal of Information Technology & Politics*, 8(1), pp.1–20.

van der Ark, L.A., Croon, M.A. and Sijtsma, K. (2008). Mokken Scale Analysis for Dichotomous Items Using Marginal Models. *Psychometrika*, 73(2), pp.183–208.

Walgrave, S., Nuytemans, M., and Pepermans, K. (2009). Voting Aid Applications and the Effect of Statement Selection. *West European Politics*, 32(6), pp. 1161-1180.

Watson, S. and Elliot, M. (2013). Entropy Balancing: A Maximum-Entropy Reweighting Scheme to Adjust for Coverage Error, *CCSR Paper*, at <www.ccsr.ac.uk/documents/Watson_Elliot.pdf>, accessed 1 August 2014.

Wheatley, J. (2012). Using VAAs to explore the dimensionality of the policy space: experiments from

Brazil, Peru, Scotland and Cyprus. *International Journal of Electronic Governance*, 5(3/4): pp. 318-348.

Wheatley, J., Carman, C., Mendez, F., and Mitchell, J. (2012). The dimensionality of the Scottish Political Space: Results from an Experiment on the 2011 Holyrood Elections. *Party Politics* (advance online publication September 30, 2012, doi: 10.1177/1354068812458614).

Table 1: Comparison of EUVox Sample Vote Intention and the Popular Vote

Party or Bloc	EUVox (%)	Election Result (%)
PP	5.51%	26.09%
PSOE	9.92%	23.01%
La Izquierda Plural	22.30%	10.03%
Podemos	19.89%	7.98%
UPyD	7.77%	6.51%
Coalición por Europa	1.75%	5.42%
L' Esquerra pel Dret a Decidir	3.76%	4.01%
Ciudadanos	5.08%	3.16%
Los Pueblos Deciden	3.47%	2.08%
Primavera Europea	9.76%	1.92%
VOX	2.78%	1.57%
Others	8.00%	8.22%

Table 2a: Sample characteristics prior to weighting

	Reference Sample (EES)		Non-random Sample (EUVox)		Difference	
	mean	standard dev.	mean	standard dev	mean	standard dev
Gender	1.458		1.707		-0.249	
Age	45.988	15.363	33.446	12.657	12.543	2.706
Education	3.124	1.764	4.377	1.209	-1.252	0.555
Political interest	2.617	0.862	1.628	0.754	0.989	0.107
Left-right self-placement	4.794	2.746	3.203	2.607	1.591	0.139

Note: Gender is represented by a dichotomous variable (1=female, 2=male); Education consistes of a six-point scale (1 - Did not complete Primary school; 2 completed Primary school; 3 - Completed high school; 4 - Technical/Vocational education; 5 - University degree; 6 - Postgraduate education); Political interest consists of a four-point scale (1 - Very interested; 2 - Somewhat interested; 3 - Not very interested; 4 - Not at all interested); left-right self placement is on a scale from (0 – far left to 10 – far right).

Table 2b: Sample characteristics after raking

	Reference Sample (EES)		Weighted Sample (EUVox)		Difference	
	mean	standard dev.	mean	standard dev	mean	standard dev
Gender	1.458		1.458		0.000	
Age	45.988	15.363	45.503	15.341	0.486	0.021
Education	3.124	1.764	3.124	1.763	0.000	0.001
Political interest	2.617	0.862	2.617	0.861	0.000	0.000
Left-right self-placement	4.794	2.746	4.794	2.745	0.000	0.001

Note: Gender is represented by a dichotomous variable (1=female, 2=male); Education consistes of a six-point scale (1 - Did not complete Primary school; 2 completed Primary school; 3 - Completed high school; 4 - Technical/Vocational education; 5 - University degree; 6 - Postgraduate education); Political interest consists of a four-point scale (1 - Very interested; 2 - Somewhat interested; 3 - Not very interested; 4 - Not at all interested); left-right self placement is on a scale from (0 – far left to 10 – far right).

Table 2c: Sample characteristics after PSA weighting (first moment only)

	Reference Sample (EES)		Weighted Sample (EUVox)		Difference	
	mean	standard dev.	mean	standard dev	mean	standard dev
Gender	1.458		1.448		0.010	
Age	45.988	15.363	46.563	17.305	-0.574	-1.943
Education	3.124	1.764	3.146	1.480	-0.022	0.284
Political interest	2.617	0.862	2.663	1.013	-0.046	-0.151
Left-right self-placement	4.794	2.746	4.878	2.924	-0.084	-0.178

Note: Gender is represented by a dichotomous variable (1=female, 2=male); Education consistes of a six-point scale (1 - Did not complete Primary school; 2 completed Primary school; 3 - Completed high school; 4 - Technical/Vocational education; 5 - University degree; 6 - Postgraduate education); Political interest consists of a four-point scale (1 - Very interested; 2 - Somewhat interested; 3 - Not very interested; 4 - Not at all interested); left-right self placement is on a scale from (0 – far left to 10 – far right).

Table 2d: Sample characteristics after PSA weighting (including first three moments and interaction terms)

	Reference Sample (EES)		Weighted Sample (EUVox)		Difference	
	mean	standard dev.	mean	standard dev	mean	standard dev
Gender	1.458		1.448		0.009	
Age	45.988	15.363	46.747	15.694	-0.758	-0.332
Education	3.124	1.764	3.081	1.768	0.044	-0.004
Political interest	2.617	0.862	2.650	0.874	-0.033	-0.012
Left-right self-placement	4.794	2.746	4.875	2.725	-0.081	0.022

Note: Gender is represented by a dichotomous variable (1=female, 2=male); Education consistes of a six-point scale (1 - Did not complete Primary school; 2 completed Primary school; 3 - Completed high school; 4 - Technical/Vocational education; 5 - University degree; 6 - Postgraduate education); Political interest consists of a four-point scale (1 - Very interested; 2 - Somewhat interested; 3 - Not very interested; 4 - Not at all interested); left-right self placement is on a scale from (0 – far left to 10 – far right).

Table 2e: Sample characteristics after entropy balancing (with reference sample unweighted)

	Reference Sample (EES)		Weighted Sample (EUVox)		Difference	
	mean	standard dev.	mean	standard dev	mean	standard dev
Gender	1.458		1.458		0.000	
Age	45.988	15.363	45.988	15.356	0.001	0.007
Education	3.124	1.764	3.124	1.763	0.000	0.001
Political interest	2.617	0.862	2.617	0.861	0.000	0.001
Left-right self-placement	4.794	2.746	4.794	2.745	0.000	0.001

Note: Gender is represented by a dichotomous variable (1=female, 2=male); Education consistes of a six-point scale (1 - Did not complete Primary school; 2 completed Primary school; 3 - Completed high school; 4 - Technical/Vocational education; 5 - University degree; 6 - Postgraduate education); Political interest consists of a four-point scale (1 - Very interested; 2 - Somewhat interested; 3 - Not very interested; 4 - Not at all interested); left-right self placement is on a scale from (0 – far left to 10 – far right).

Table 2f: Sample characteristics after entropy balancing (with reference sample weighted)

	Reference Sample (EES)		Weighted Sample (EUVox)		Difference	
	mean	standard dev.	mean	standard dev	mean	standard dev
Gender	1.489		1.489		0.000	
Age	46.979	17.300	46.979	17.300	0.000	0.000
Education	1.693	1.257	1.693	1.257	0.000	0.000
Political interest	2.867	0.842	2.867	0.842	0.000	0.000
Left-right self-placement	4.955	2.924	4.954	2.924	0.000	0.000

Note: Gender is represented by a dichotomous variable (1=female, 2=male); Education consistes of a six-point scale (1 - Did not complete Primary school; 2 completed Primary school; 3 - Completed high school; 4 - Technical/Vocational education; 5 - University degree; 6 - Postgraduate education); Political interest consists of a four-point scale (1 - Very interested; 2 - Somewhat interested; 3 - Not very interested; 4 - Not at all interested); left-right self placement is on a scale from (0 – far left to 10 – far right).

Table 3: Sample Characteristics for Vote Intention

Party	Final Vote (%)	EUVox unweighted %	EUVox (raking) %	EUVox (PSA, standard) %	EUVox (PSA, with higher moments) %	EUVox (EB) %	EUVox (EB, with EES weights) %
PP	26.09	5.14 (4.94, 5.34)	14.77 (11.41, 18.92)	13.90 (11.87, 16.22)	17.38 (13.56, 22.00)	16.62 (13.18, 20.76)	22.33 (14.37, 33.01)
PSOE	23.01	9.47 (9.21, 9.74)	11.26 (9.55, 13.23)	12.59 (11.32, 13.98)	9.92 (8.48, 11.57)	10.36 (8.91, 12.02)	10.17 (7.31, 13.98)
La Izquierda Plural	10.03	23.19 (22.82, 23.57)	13.19 (11.85, 14.65)	14.32 (13.30, 15.42)	13.28 (11.76, 14.96)	13.67 (12.16, 15.32)	12.65 (9.38, 16.86)
Podemos	7.98	19.86 (19.51, 20.22)	17.17 (14.35, 20.42)	15.73 (14.46, 17.09)	15.32 (13.33, 17.54)	15.60 (13.71, 17.71)	15.62 (11.86, 20.29)
UPyD	6.51	7.63 (7.40, 7.88)	10.09 (9.00, 11.30)	8.86 (7.94, 9.87)	9.83 (8.42, 11.45)	9.82 (8.50, 11.33)	7.78 (5.05, 11.80)
Coalición por Europa	5.42	1.80 (1.69, 1.93)	1.53 (1.24, 1.88)	1.71 (1.34, 2.18)	1.51 (1.19, 1.90)	1.50 (1.20, 1.86)	0.92 (0.57, 1.47)
L' Esquerra pel Dret a Decidir	4.01	3.92 (3.75, 4.09)	2.14 (1.82, 2.52)	2.46 (2.17, 2.80)	2.15 (1.79, 2.58)	2.21 (1.85, 2.63)	1.97 (1.24, 3.12)
Ciudadans	3.16	5.09 (4.90, 5.29)	8.39 (6.59, 10.61)	7.95 (6.48, 9.72)	8.18 (6.54, 10.19)	8.03 (6.49, 9.91)	6.45 (3.94, 10.38)
Los Pueblos Deciden	2.08	3.62 (3.45, 3.79)	1.43 (0.90, 2.26)	1.30 (1.12, 1.50)	1.21 (0.90, 1.63)	1.23 (0.90, 1.68)	1.51 (0.63, 3.57)
Primavera Europea	1.92	10.15 (9.88, 10.42)	5.90 (4.87, 7.13)	6.69 (5.36, 8.31)	6.46 (4.82, 8.60)	6.35 (5.07, 7.94)	4.38 (2.75, 6.89)
VOX	1.57	2.76 (2.61, 2.91)	4.83 (3.88, 6.01)	5.74 (4.82, 6.82)	5.92 (4.21, 8.26)	5.72 (4.40, 7.42)	5.29 (3.45, 8.04)
RED	0.67	1.01 (0.92, 1.10)	1.23 (0.83, 1.82)	1.12 (0.80, 1.58)	1.24 (0.81, 1.91)	1.27 (0.83, 1.95)	1.79 (0.73, 4.29)
Others	7.55	6.37 (6.15, 6.59)	8.07 (6.26, 10.34)	7.62 (6.13, 9.44)	7.59 (5.75, 9.97)	7.59 (5.87, 9.77)	9.14 (5.95, 13.79)

Table 4: Sample Characteristics for Attitudes towards immigrants

Party	EES unweighted %	EES weighted %	EUVox unweighted %	EUVox (raking) %	EUVox (PSA, standard) %	EUVox (PSA, with higher moments) %	EUVox (EB) %	EUVox (EB, with EES weights) %
Completely Agree	32.87 (29.94, 35.95)	38.27 (33.49, 43.29)	17.02 (16.77, 17.26)	34.26 (31.95, 36.65)	32.69 (30.71, 34.73)	34.54 (31.91, 37.27)	33.99 (31.70, 36.35)	42.34 (37.53, 47.31)
Agree	50.74 (47.54, 53.94)	48.16 (43.22, 53.14)	33.64 (33.33, 33.95)	38.94 (36.79, 41.13)	39.80 (37.68, 41.97)	39.33 (36.62, 42.11)	39.00 (36.79, 41.26)	34.81 (30.43, 39.46)
Neither Agree nor Disagree	8.30 (6.69, 10.24)	7.03 (4.85, 10.08)	22.33 (22.06, 22.61)	13.77 (12.46, 15.21)	13.71 (12.91, 14.54)	12.89 (11.82, 14.04)	13.36 (12.35, 14.45)	10.03 (8.13, 12.32)
Disagree	7.02 (5.55, 8.84)	6.23 (4.23, 9.09)	19.77 (19.51, 20.04)	9.98 (9.07, 10.96)	10.60 (9.97, 11.26)	10.25 (9.09, 11.53)	10.53 (9.45, 11.71)	9.25 (7.18, 11.84)
Completely Disagree	1.07 (0.57, 1.97)	0.31 (0.15, 0.61)	7.24 (7.07, 7.41)	3.05 (2.73, 3.41)	3.20 (2.93, 3.50)	2.99 (2.58, 3.48	3.12 (2.68, 3.61)	3.57 (2.52, 5.01)

Table 5. Characteristics of Sub-Samples of the England Dataset

Variable	Whole sample (N=70,201)	Daily Mail Sample (N=32,877)	Guardian Sample (N=14,190)
Year of Birth (mean)	1974	1969	1976
Year of Birth (median)	1976	1969	1980
Higher Education %	63.2	49.3	77.1
% "Very interested" in politics	33.0	30.7	34.9
Left-Right self-placement (mean)	4.82	6.22	3.84
Liberal-Conservative self-placement (mean)	4.01	6.07	2.84
Vote intention %	UK Independence Party – 35.3 Labour Party – 21.0 Conservative Party – 15.7 Liberal Democrats – 8.0 Green Party – 18.4 Others – 1.6	UK Independence Party – 60.8 Labour Party – 10.4 Conservative Party – 20.4 Liberal Democrats – 3.7 Green Party – 3.3 Others – 1.3	UK Independence Party – 13.1 Labour Party – 32.3 Conservative Party – 9.7 Liberal Democrats – 10.8 Green Party – 32.2 Others – 1.9

Table 6. Goodness-of-fit indices (CFA) and H coefficients (Mokken Scale Analysis), whole sample (N=70,201)

		Scales		
Item	**Issue Statement**	**Dim 1**	**Dim 2**	**Dim 3**
1	The United Kingdom should never adopt the Euro	0.543		
3	The right of EU citizens to work in the United Kingdom should be restricted	0.639		
6	Overall, EU membership has been a bad thing for the United Kingdom	0.647		
10	The United Kingdom should hold an in or out referendum on EU membership as soon as possible.	0.572		
11	Free market competition makes the health care system function better		0.418	
12	The number of public sector employees should be reduced		0.484	
13	The state should intervene as little as possible in the economy		0.439	
15	Cutting government spending is a good way to solve the economic crisis		0.468	
16	It should be easy for companies to fire people		0.400	
19	The top rate of income tax should be reduced further.		0.433	
21	Immigrants must adapt to the values and culture of the United Kingdom	0.599		
22	Restrictions on citizen privacy are acceptable in order to combat crime			0.555
23	To maintain public order, governments should be able to restrict demonstrations			0.555
25	Same sex couples should enjoy the same rights as heterosexual couples to marry	0.498		
28	Islam is a threat to the values of the United Kingdom	0.594		
29	The United Kingdom should welcome a larger number of asylum seekers from war-torn countries.	0.581		
Overall scalability coefficient (H) for each scale		0.585	0.441	0.555
Comparative Fit Index (CFI)			0.968	
Tucker-Lewis Index (TFI)			0.962	
Root Mean Square Error of Approximation (RMSEA)			0.078	

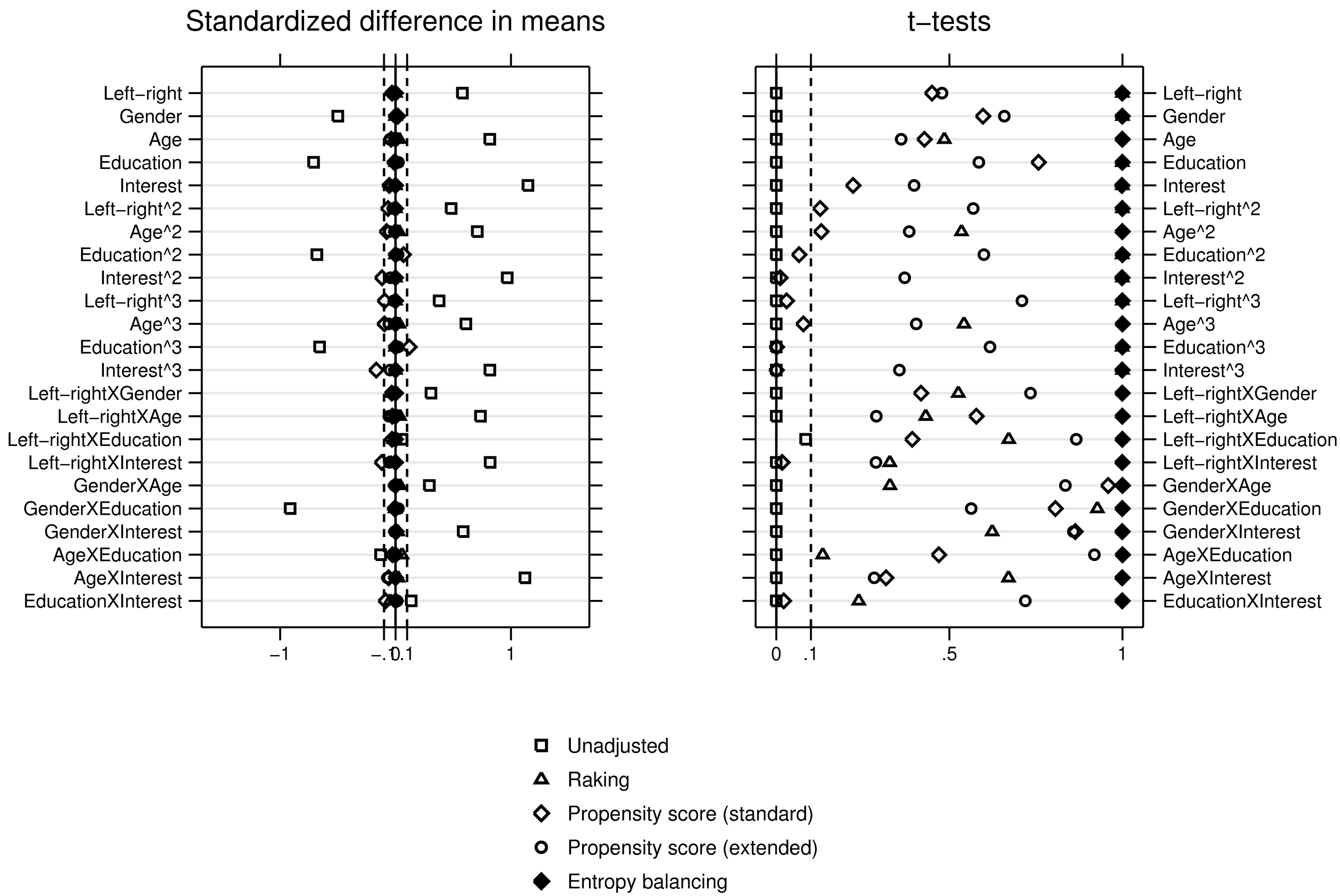
Table 7. Goodness-of-fit indices (CFA) and H coefficients (Mokken Scale Analysis), Daily Mail sample (N=32,877)

		Scales		
Item	**Issue Statement**	**Dim 1**	**Dim 2**	**Dim 3**
1	The United Kingdom should never adopt the Euro	0.460		
3	The right of EU citizens to work in the United Kingdom should be restricted	0.549		
6	Overall, EU membership has been a bad thing for the United Kingdom	0.569		
10	The United Kingdom should hold an in or out referendum on EU membership as soon as possible.	0.508		
11	Free market competition makes the health care system function better		0.308	
12	The number of public sector employees should be reduced		0.393	
13	The state should intervene as little as possible in the economy		0.349	
15	Cutting government spending is a good way to solve the economic crisis		0.367	
16	It should be easy for companies to fire people		0.324	
19	The top rate of income tax should be reduced further.		0.329	
21	Immigrants must adapt to the values and culture of the United Kingdom	0.512		
22	Restrictions on citizen privacy are acceptable in order to combat crime			0.513
23	To maintain public order, governments should be able to restrict demonstrations			0.513
25	Same sex couples should enjoy the same rights as heterosexual couples to marry	0.394		
28	Islam is a threat to the values of the United Kingdom	0.500		
29	The United Kingdom should welcome a larger number of asylum seekers from war-torn countries.	0.472		
Overall scalability coefficient (H) for each scale		0.496	0.345	0.513
Comparative Fit Index (CFI)			0.950	
Tucker-Lewis Index (TFI)			0.941	
Root Mean Square Error of Approximation (RMSEA)			0.075	

Table 8. Goodness-of-fit indices (CFA) and H coefficients (Mokken Scale Analysis), Guardian sample (N=14,190)

		Scales		
Item	**Issue Statement**	**Dim 1**	**Dim 2**	**Dim 3**
1	The United Kingdom should never adopt the Euro	0.447		
3	The right of EU citizens to work in the United Kingdom should be restricted	0.555		
6	Overall, EU membership has been a bad thing for the United Kingdom	0.556		
10	The United Kingdom should hold an in or out referendum on EU membership as soon as possible.	0.455		
11	Free market competition makes the health care system function better		0.473	
12	The number of public sector employees should be reduced		0.487	
13	The state should intervene as little as possible in the economy		0.474	
15	Cutting government spending is a good way to solve the economic crisis		0.494	
16	It should be easy for companies to fire people		0.421	
19	The top rate of income tax should be reduced further.		0.462	
21	Immigrants must adapt to the values and culture of the United Kingdom	0.486		
22	Restrictions on citizen privacy are acceptable in order to combat crime			0.572
23	To maintain public order, governments should be able to restrict demonstrations			0.572
25	Same sex couples should enjoy the same rights as heterosexual couples to marry	0.416		
28	Islam is a threat to the values of the United Kingdom	0.499		
29	The United Kingdom should welcome a larger number of asylum seekers from war-torn countries.	0.496		
Overall scalability coefficient (H) for each scale		0.490	0.469	0.572
Comparative Fit Index (CFI)			0.951	
Tucker-Lewis Index (TFI)			0.941	
Root Mean Square Error of Approximation (RMSEA)			0.080	

Figure 1. Plots of Standardized Bias and T-tests



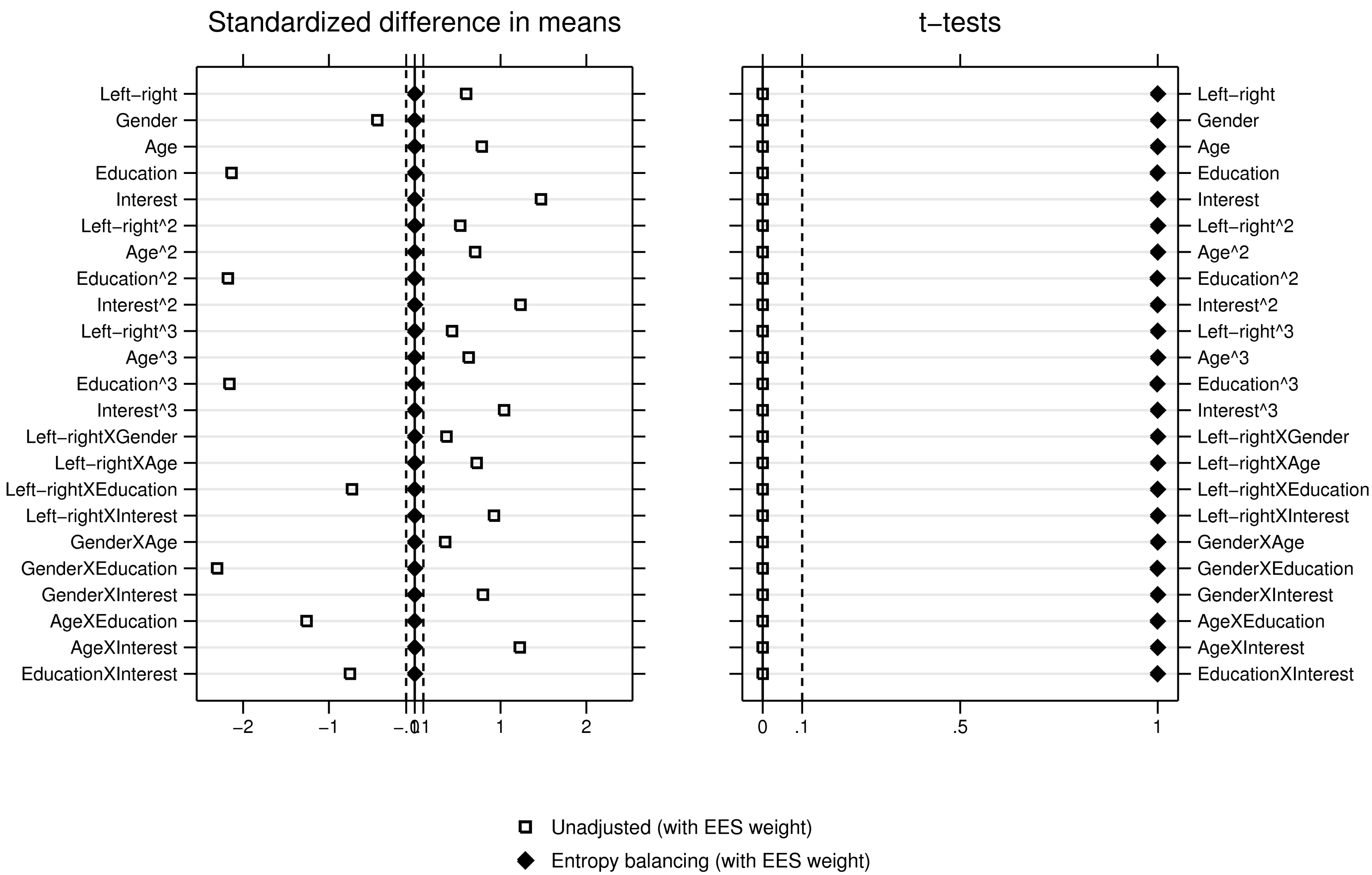

Standardized difference in means
t−tests
Left−right
Gender
Age
Education
Interest
Left−right^2
Age^2
Education^2
Interest^2
Left−right^3
Age^3
Education^3
Interest^3
Left−rightXGender
Left−rightXAge
Left−rightXEducation
Left−rightXInterest
GenderXAge
GenderXEducation
GenderXInterest
AgeXEducation
AgeXInterest
EducationXInterest
−2
−1
−.1
0
.1
1
2
0
.1
.5
1
Unadjusted (with EES weight)
Entropy balancing (with EES weight)

Figure 2. Scree Plots