The Q-Block Method and an Alternative: Benchmarking the Australian Discourse of Democracy

Simon Niemeyer
Australian National University

Abstract. *This paper uses a set of overlapping statements between a 1993 study on Australian discourses of democracy performed by Dryzek (1994) and a more recent Q study in association with the Citizen's Parliament held in 2009 to explore the relative merits of using the Q-block method, proposed by Talbott (1963; reprinted in this issue) and a simpler 'z-score method' for indexing the relationship between individuals and a pre-existing set of factors established by the Dryzek study. The results reveal a stronger correlation with the original study factor loadings using the z-score method than using the Q-block method, which tended to systematically underestimate associations with factors. Although the Q-block method remains the easier of the two methods for obtaining data, the z-score method produces more accurate results as a proxy for individual Q-sort factor loadings.*

Introduction

Q methodology is a powerful tool for exploring subjectivity in relation to a particular issue or topic, but it is best suited to small numbers of participants, not least because of the demanding process involved in obtaining a Q sort. Most Q studies are primarily concerned with identifying factors describing different orientations around an issue or topic, but in some cases it is desirable to investigate the relative prevalence of factors within a sampled population—or the indexing of Q typologies, to use Talbott's (1963) terminology. In such cases it may either be impractical to implement a Q study involving a large number of Q sorts, or in the case of a longitudinal study looking at changes to prevalence over time, the content of the Q statement set may vary slightly due to a range of conflicting research goals.

One such instance of the latter problem arose in relation to a study of Australia's first Citizens' Parliament (CP) on Australian democracy in 2009 (see http://www.citizensparliament.org.au). The study provided an opportunity to compare the initial Q sorts provided by CP

Contact author: simon.niemeyer@anu.edu.au
Operant Subjectivity: The International Journal of Q Methodology, 2010, 34(1): 59–84

participants to the prevalence of the Australian discourses of democracy identified by John Dryzek in his 1993 Q study (Dryzek, 1994). It was not possible to simply re-use Dryzek's original Q-statement set, but it was possible to use a subsample of those statements for the CP study.

In order to compare the CP and Dryzek studies, what was needed was a method for using the sample of statements common to both studies to benchmark participants against the original Dryzek factors. The Q-block method provided a tantalizingly simple, but ultimately inferior, solution. The method itself is intuitively appealing, permitting the use of a subsample of Q statements in a shortened 'Q-block' survey, the results of which can then be used to calculate associations with factors or derived factor loadings. The premise is relatively straightforward. However, the method itself is underdeveloped. The only known reference to it is a 1963 conference paper presented by Albert Talbott.

This paper attempts to advance the development of methods for benchmarking Q typologies. It explores the viability of the Q-block method compared to an alternative approach based on the use of factor *z*-scores. These two methods are used to benchmark the CP participants' Q sorts to Dryzek's 1993 discourses of Australian democracy. A comparison of the two approaches is then obtained using the original Dryzek data to perform exactly the same analysis as for the CP data to see which of the two approaches produces proxy factor loadings most closely correlated to those from the original study. The focus on factor loadings is slightly different from the indexing approach used by Talbott, which does not generate a proxy factor loading per se, but an index that associates individuals with factors. However, even if association is the focus for a particular researcher, investigating proxy factor loadings will still be worthwhile because factor loadings are the measure used in Q methodology to associate individuals with factors. So it is worth focussing on this dimension, not least, as will be seen, because the results of the analysis raise questions about how associations with factors should actually be determined.

The paper begins by briefly revisiting the results of Dryzek's (1994) study, describing the four factors relating to Australian discourses of democracy. Then the two methods for benchmarking the CP study data against these factors are implemented, followed by a testing of the reliability of the two approaches. The concluding discussion then considers the relative merits of the two approaches.

Original Study: Australian Discourses of Democracy

In his 1993 study, Dryzek (1994) used Q methodology to reconstruct discourses capable of describing how Australians encounter their system of democracy. The study involved 60 participants from five

Australian sites (Brisbane, Dalby, Sydney, Canberra, and Geelong). The statements themselves were selected from a larger sample "pertaining to the collective construction, distribution, application, and limitation of political authority" (Dryzek, 1994, p. 223) drawn from various newspaper sources, political manifestos and works on Australian politics, with the final 64 statements selected to populate a 4x4 matrix of discourse elements (relating to ontology, agency, motivations or natural relationships) and type of claim (definitive, designative, evaluative and advocative).

In the original Dryzek study, participants sorted the 64 statements across 13 piles depending on whether they agreed or disagreed with them on a +6 (most agree) to -6 (most disagree) scale. Centroid factor analysis followed by varimax rotation produced four factors.

Australian Discourses of Democracy

The original study factor scores are provided in the Appendix. The following section provides a brief description of each factor.

Resigned Acceptance (Discourse A). Discourse A is associated with a cynical belief that Australia does not have a true democracy. It is an inflexible and unrepresentative system. Money controls politics. The rich are able to manipulate the system to their advantage. The country is run by a cabal in Canberra, who represents a minority of powerful interests and tell the rest of the country what to do. The rest of the citizens are poorly consulted, if at all. Their votes are viewed as ineffective in making a difference, particularly given the poor choices from the machinations of party politics. This is not a great system, but we are stuck with it and probably cannot achieve much better.

Inclusive Republicanism (Discourse B). As for Discourse A, there is a strong concern regarding particular facets of Australian democracy, but there is also a belief that things can change. Of particular concern is the lack of representation of women in parliament, followed by the brainwashing of citizens by television news and advertising and the influence of money. Unlike the resignation associated with Discourse A, this is a reformist discourse. There is a belief that our democracy should be improved to achieve a more humane and equitable society, in which citizens have a sense that they can make a difference. Ideally, improvements would lead to a society that avoids the excessive influence of any particular group, be it majority or minority—one that is equitable and humane. Such values are more important than the pursuit of economic rationality, which so pervades the current system. One important element of reform is a move toward republicanism.

Right-Minded Democracy (Discourse C). Australian democracy may not be held in high esteem in Discourse C but it is nonetheless promulgated by 'true' believers in the notion of democracy. But truth here refers to a

particular democratic discourse that places great emphasis on the actions of, and expectations ascribed to, political actors in a democracy that is primarily about leadership by those elected by the majority. Democracy is only broken in practice, due to the behaviour of its main actors. Leaders might have failed, but could restore democratic potential if they are moral in their actions—in accordance with the Bible. Citizens are too easily led astray by the dictates of sound-bite television and should be more active members of the community in which too few are influencing political outcomes. Although leadership-oriented, this position is not averse to democratic discourse. But this is not deliberative democracy, per se. Rather, there is a sense that democracy should encourage different views in the same sense that a proselytizer might encourage dialogue. There is a hope that discourse will ultimately produce a moral leadership that can be entrusted to maintain peace and justice.

Anxious Egalitarianism (Discourse D). Australia is a great democratic country, at least in principle, but it is not without its problems. Too many small powerful groups are getting too much of a say at the cost of the rest of us when what we really need is pluralism in our democratic discourse. The yawning gap between the real and the ideal in Australian democracy yields a cynical outlook toward those political actors that distort our otherwise great system. And it does not help that too many of us blunder into the polling booth without finding out more about the issue that we are voting on.

Factor Abundance

The average factor loadings for the four factors obtained by the original Dryzek study are 0.26, 0.23, 0.12, and 0.18 respectively for factors A, B, C and D. Dryzek sets the level of association with a factor loading with a factor at 0.45, and at this level there are 16, 13, 4 and 5 individuals uniquely loaded on each of the factors respectively.

The objectives of the remainder of the paper are to determine the extent to which the abundance of the factors for the CP participants differed from the original study using two methods for indexing typologies and to test the accuracy of the methods.

Benchmarking the Original Study Discourses

The Citizens' Parliament comprised 150 Australians drawn randomly from over 3,000 responses to an invitation posted to 7,000 randomly selected addresses. All respondents to the initial invitation were asked to perform an online Q sort as part of the recruitment. It is these Q sorts that were used to benchmark against the original Dryzek study.

Using the CP study to compare to Dryzek's study posed considerable challenges. The CP involved a multi-stage Q study in its own right, with a

slightly different, if strongly overlapping set of issues, from the original Dryzek study. It was not appropriate to simply re-use the Q set from the original study. But the overlapping issues presented the possibility to use a sample of statements from the original Dryzek study, combined with additional statements developed specifically for the CP study.

Selection of these statements was done in two ways. Firstly, the Q-block method described by Talbott (1963) was used to draw representative statements from the original statement set. The basic method involved establishing a set of Q blocks containing representative statements from each of the Q factors defined by Dryzek, ranking the statements within a block, and aggregating the ranks to determine how strongly an individual concords with each factor.

The number of statements in each block is equal to the number of factors being indexed. A selection of statements for each of the sets is chosen so that each factor is 'represented' by a statement such that the factor score (or, more accurately, the *z*-score) is relatively high at about the same level for each of the factors represented. Secondly, the *z*-score should be substantially higher for the factor represented than for the remainder of the factors. According to Talbott (1963) there should be one standard error separation between the *z*-score of the factor being represented by the statement and the *z*-scores for the remaining factors for that statement *and* the *z*-scores for the remaining statements in the same block. Looking at the first four statements comprising Block 1 in Table 1 (p. 65 below), this means that the *z*-score for Factor A for the first statement should be higher than the *z*-scores in the same row and the same column, and so on for the remaining diagonal scores in the table for that block.

In practice, meeting the criteria set by Talbott was impossible to achieve for the original-study factors. For example, it can be seen from the first row in Table 1 that the difference in *z* score between Factor A and Factor B is only 0.68. The rationale for the standard-error criterion is not provided by Talbott, but adhering to it would have rendered this Q-block exercise futile.

Notwithstanding the standard-error criterion, the actual number of blocks that should be used in a Q-block study is also not clear from Talbott's paper. His discussion and example regarding how to construct Q blocks (Talbott, this issue, pp. 7–8) seem to suggest that four blocks can be used to benchmark four factors, requiring 16 statements, with each block containing a representative statement from each factor that has a substantially higher *z*-score to the other factors. But later in the paper, Talbott goes on to describe how 24 Q blocks were constructed comprising 13 positive and 11 negative blocks (Talbott, this issue, p. 9), where negative blocks contain representative statements from each

factor that are strongly *negatively* loaded and substantially different from the other factors.

Just how Talbott did this using 57 statements is difficult to discern from his paper. Not all 24 blocks are reproduced. Given that, each block requires a representative statement from each of four factors and that 24 blocks covering four factors would requires 96 statements, it is clear that statements were used more than once across the entire range of blocks.

In light of both the constraints in designing blocks to fit Talbott's criteria and questions about how many blocks should actually be developed for a study, it was decided to use a number of strategies for constructing Q blocks. The first strategy involved constructing blocks that met as closely as possible the criteria described by Talbot to achieve the minimum requirement of four positive blocks using 16 unique statements. These are the first four blocks shown in Table 1.

The second strategy involved constructing a number of negative blocks (where the statement has a strongly negative *z*-score for the factor that it represents). Two blocks—numbers 5 & 6—were constructed this way, involving some repetition of statements used in the first four blocks. The third strategy involved constructing a further four blocks (7–10) using a more relaxed version of Talbott's specified requirements, where the minimum separation between *z*-scores was 0.5 in addition to the repeated use of statements from one block in different combinations. Together, these strategies used 24 of the statements from the original study that were incorporated in the Q-statement set for the CP study.

A further seven statements were selected from the Q set from the original study because they were particularly relevant to the new context on their own merit, even though the associated *z*-scores did not fit either Talbott's criteria or the more relaxed version. These statements were not used as part of the Q-block analysis, but they were used in the *z*-score analysis. Together this resulted in 31 statements common to both studies in addition to 17 that are specific to the CP study.

Obtaining the Data

Obtaining the data from participants to be used for comparison with the data from the original Dryzek study also required some careful methodological consideration. According to Talbott's method, implementation of a Q-block survey involves obtaining rankings of statements within each of the n statement sets, which can then be aggregated to obtain a measure of correlation with each factor. Although he recognises that other methods are possible, the instrument used by Talbott involves ranking each of the statements (representative of the factors) within each of the Q blocks.

Table 1: Q-Block Design

Block	Statement (Original number, CP number)	Z-scores (factor) A	B	C	D
1	In Australia the rich have virtually unlimited access to the legal system and the capacity to use it to achieve their own ends. (21, 9)	2.18	1.50	-0.98	-0.42
	Until women are equally represented in Australian parliaments, we will not have a genuine representative democracy. (16, 25)	-1.11	2.20	-0.88	-2.04
	If Australia is to recover from the horrible mess we are in at present, we will need leaders of high principles, who will govern with honesty, justice, and righteousness, fearlessly putting into practice the principles of the Bible. (58, 23)	-1.68	-1.00	2.21	-1.79
	We are a democratic country. (19, 4)	-0.18	-0.35	-0.71	3.19
2	There's a dollar democracy that runs through our supposed democracy. (18, 3)	2.07	0.26	-0.75	0.80
	A lot of people may be politically ignorant, but they believe what they see and hear on television, and that makes them vote the way they do. They have been brainwashed by the clever advertising people. (32, 28)	0.73	1.78	0.99	0.58
	Governments can talk a great game about 'democracy'. They fall down, in practice, because they cannot accept any sharing of the power to influence decisions with those who might have different motives and ideas. (42, 20)	0.44	0.35	1.68	-0.52
	Sometimes I get a bit worried that there's a lot of small minority groups that tend to get a lot of the say. (39, 12)	0.15	-0.47	1.17	1.68
3	I suppose we do have an upper class and a lower class. (29, 26)	1.89	0.78	-1.12	1.20
	Our community and nation should invoke a spirit that you are important, that you do have a way of influencing things. (59, 24)	0.65	1.57	0.83	0.67
	The true believers are about decency in life, about a capacity to debate the issues and to receive different views. (10, 15)	0.09	0.72	1.65	0.92
	I don't want to be in a place where only a minority's viewpoint gets the right to say what's going to happen. (56, 14)	1.02	0.13	0.93	1.52

4	I'm always cynical about government processes. (27, 19)	1.64	0.31	-1.17	-0.13
	Concluding that some races/cultures are better than others must be avoided. (63, 31)	0.53	1.33	-0.08	0.85
	Free markets work because individual people, cooperating peacefully and voluntarily through markets, can achieve much that politicians and bureaucrats cannot achieve using compulsion and direction. (45, 29)	-0.13	-1.12	1.28	0.40
	The only time we are all equal is when we vote in the ballot box. (31, 27)	-1.20	0.31	-0.73	1.27
5	The only time we are all equal is when we vote in the ballot box. (31, 27)	-1.20	0.31	-0.73	1.27
	Republicanism is a smokescreen for democratic treachery, incompetence and lunacy. (12, 17)	-0.74	-2.35	0.85	-0.97
	I'm always cynical about government processes. (27, 19)	1.64	0.31	-1.17	-0.13
	Until women are equally represented in Australian parliaments, we will not have a genuine representative democracy. (16, 25)	-1.11	2.20	-0.88	-2.04
6	Our political system's strength is that it can change to meet changing circumstances. (35, 7)	-1.34	-0.69	0.64	-0.32
	I don't think there's anything wrong with people voting for their self-interest. (43, 21)	0.84	-1.13	-0.36	0.99
	Our government is attempting to mould our society to the needs of a profit-oriented market. (26, 18)	0.01	0.13	-1.16	0.90
	The last thing we need is a Canberra 'Big Brother' telling us what we should read and hear. (50, 8)	0.76	0.45	0.52	-0.89
7	If you want your vote to count for anything, then you need to not vote in ignorance, you need to know something about what you're being offered. (7, 2)	1.71	1.50	0.89	1.48
	Low feminist participation rates compound the domination of the public realm by males. (48, 30)	-0.59	0.78	-0.07	-0.71
	You don't have to wait three years to say 'I'm not happy with the job you did.' You can apply pressure, and enough people writing to the local member can turn the situation around. (22, 10)	-0.82	0.17	1.10	-0.76
	I don't want to be in a place where only a minority's viewpoint gets the right to say what's going to happen. (56, 14)	1.02	0.13	0.93	1.52

Block	Statement (Original number, CP number)	Z-scores (factor) A	B	C	D
8	There's a dollar democracy that runs through our supposed democracy. (18, 3)	2.07	0.26	-0.75	0.80
	Until women are equally represented in Australian parliaments, we will not have a genuine representative democracy. (16, 25)	-1.11	2.20	-0.88	-2.04
	The true believers are about decency in life, about a capacity to debate the issues and to receive different views. (10, 15)	0.09	0.72	1.65	0.92
	The only time we are all equal is when we vote in the ballot box. (31, 27)	-1.20	0.31	-0.73	1.27
9	In Australia the rich have virtually unlimited access to the legal system and the capacity to use it to achieve their own ends. (21, 9)	2.18	1.50	-0.98	-0.42
	A lot of people may be politically ignorant, but they believe what they see and hear on television, and that makes them vote the way they do. They have been brainwashed by the clever advertising people. (32, 28)	0.73	1.78	0.99	0.58
	Free markets work because individual people, cooperating peacefully and voluntarily through markets, can achieve much that politicians and bureaucrats cannot achieve using compulsion and direction. (45, 29)	-0.13	-1.12	1.28	0.40
	We are a democratic country. (19, 4)	-0.18	-0.35	-0.71	3.19
10	I suppose we do have an upper class and a lower class. (29, 26)	1.89	0.78	-1.12	1.20
	Our community and nation should invoke a spirit that you are important, that you do have a way of influencing things. (59, 24)	0.65	1.57	0.83	0.67
	If Australia is to recover from the horrible mess we are in at present, we will need leaders of high principles, who will govern with honesty, justice, and righteousness, fearlessly putting into practice the principles of the Bible. (58, 23)	-1.68	-1.00	2.21	-1.79
	Sometimes I get a bit worried that there's a lot of small minority groups that tend to get a lot of the say. (39, 12)	0.15	-0.47	1.17	1.68

This particular approach was not possible for the CP study, because the Q-block component was being administered as part of a new Q study.

The actual approach that was used involved administering the CP Q study using a slightly modified set of instructions that would permit the construction of a Q-block ranking from the study data. This involved using the forced rankings for the individual statements from each Q sort and then ranking them within each block.

Because of the high probability of statements within each block ending up with a tied forced response in the Q sort, it was decided to add a step at the beginning of the Q sort where participants were asked to provide an unforced Likert response to each of the statements, where the Likert response was then used to distribute the statements between three piles, after which Q sorting proceeded according to the usual method. To facilitate the implementation of this method over the net, modifications were done to the FlashQ program (Hackert & Braehle, 2006) to substitute the process of Likert rating for the step involving the allocation of cards between 'disagree', 'middle' and 'agree' piles prior to Q sorting, with the allocation between the piles being done automatically based on the unforced response (less than -2 being allocated to the disagree pile, greater than +2 to the agree pile, with the remainder being allocated to the middle pile).

The unforced Likert rating of each statement was done using the same 11-point scale as for the forced data. The unforced response was between -5 and +5 according to level of agreement. The sorting process was done on a scale of 1 to 11 based on increasing level of agreement with the quotas for each of the 11 piles being 2,3,4,6,6,6,6,6,4,3,2.

Thus, the results obtained from the Q sorts from the CP study are different to those using the original Talbott (1963) method. In the original method, respondents directly rank statements within a block. The block statements were interspersed among a wider set of statements as part of the CP Q study. However, by obtaining unforced and forced responses to the block statements it is possible to employ a number of methods for obtaining derived factor loadings. Two of these methods are used in the following section, with the results compared.

Analysis

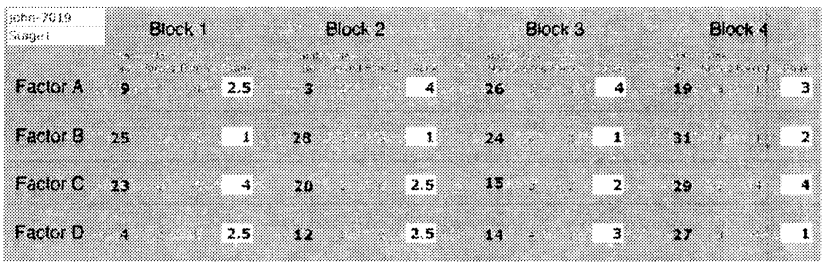
Using the raw data from the Q sorts it was possible to obtain a derived factor loading for each individual respondent using two main approaches. The first method is closely related to the block method described by Talbott (1963). The second method involves correlating the forced responses to the block statements to the factor *z*-scores from the original study.

Method 1: Block Ranking

In order to perform a block-method analysis along the lines developed by Talbott (1963) it was necessary to obtain block rankings for the four

statements in each block for every participant. As previously mentioned, this was done by sorting each individual's responses within a block by forced followed by unforced response. The actual process of block ranking was done using a database program (Filemaker Pro), where a script was written to sort forced responses within each block, followed by unforced responses, with the resulting rank being allocated in order from most to least agree. An example of how this is done is provided in Figure 1, which shows a window capture from the database program.

Figure 1. Sample Block-Ranking Calculation Based on Unforced and Forced Statement Responses



It can be seen from Figure 1 that in some cases it was not possible to obtain untied ranks using both forced and unforced data. For example, in Block 1 statement 9 and 4 are tied, with both having a forced rank of 0 and unforced rating of 3, resulting in both being given a ranking of 2.5. A similar situation has occurred in Block 2. However for Block 3 and 4, ties between forced responses could be differentiated using the unforced data. For example statements 26 and 14 both have a forced rank of 1, but 14 has a higher unforced rating, so is ranked higher in the Q block.

Associating individuals with factors using the Q-block method

The next step involves using the block ranks to derive a correlation measuring how strongly a participant is associated with each of the four factors from the original study. Two different strategies involving different numbers of blocks were used to obtain indexes (or more specifically in this case, proxy correlations) associating individuals with factors. The first involves using only the first four (positive) blocks that most closely satisfy Talbott's minimum requirements. The second involves using the extended set of 10 blocks, including the additional six blocks obtained using the methods described above.

In terms of actually calculating the index of association with a factor, the original Talbott study involved aggregating block ranks for each of the factors. So, using the example in Figure 1 where four blocks are used, the total of rank for Factor A is 13.5. Because a total rank of 16 is the lowest possible with the statements in the block, (the highest being a rank of 4), in this case it appears that john-7019 has a low association

with Factor A.

The threshold, or criterion, for associating an individual with a factor is not clearly described by Talbott. In any case a slightly different approach was used for association, where block ranks are converted into correlations as a form or proxy factor loadings. Using this method, an aggregate rank of 13.5 translates into a proxy coefficient indicating agreement with the factor, in this case being 0.25.*

Using this method the derived factor loading (Q-block method) has been calculated for all the CP participants using the first four blocks, followed by the larger set of 10 blocks. The average loadings are reproduced in Table 2, with the results from the original study shown for comparison.

Table 2. Comparison of Average Factor Loadings from the Original Study with Derived Average Factor Loadings for the CP Study using the Q-Block Method

	A	*B*	*C*	*D*
Four Blocks	-0.04	0.31	-0.28	0.01
Ten Blocks	0.01	0.24	-0.21	-0.04

Both approaches using different sets of blocks produce similar outcomes. The average correlation between the two methods is 0.86. Although this suggests that there is not much difference between the approaches using different numbers of blocks, it does not tell us how accurate either method is for benchmarking the factors.

Z-Score Correlation Method

Another approach to obtaining a derived factor loading involves using all the available statements that are common to both studies. In this method the derived factor loading is simply obtained by correlating (for each individual) the forced response for the Q-block statements in their Q sort with the *z*-scores for each of the factors from the original study.†

Given that there are seven additional statements to the 24 Q-block statements that are shared between the two studies in the expanded set of 10 blocks, it is possible to further expand the set of statements using the *z*-score method to 31. To compare the effect of the sample size

---

* Transformation to derived correlations is done using the formula:

$$\ell' = (SUMR - A)/B$$

where SUMR is the aggregation of block ranks for a factor, and A = Max/2+Min and B = (Max-Min)/2 for a maximum and minimum aggregate rank.

† In theory it is also possible to use the factor scores as well, but because the original study had 13 piles, compared to the 11 used for CP study, the factor scores are not directly comparable. Z-scores, on the other hand, do not suffer from the same problem of commensurability.

of statements, the *z*-score correlations were done using both the 16 statements from the smaller block set of four and the entire suite of 31 shared statements, with the resulting averages to the derived factor loadings shown in Table 3.

Table 3. Average (derived) Factor Loadings Using the Z-Score Method

	Factor			
Method	***A***	***B***	***C***	***D***
Z-score method (16 block statements)	0.16	0.24	-0.09	0.30
Z-score method (All shared statements	0.27	0.35	0.03	0.35

The *z*-score method is conceptually much simpler than the Q-block method. All that is needed is a ranking of statements so that the resulting array can be correlated to the *z*-scores obtained from the original study that is being benchmarked. There are a number of finer-grained considerations, concerning whether or not the individual ranks obtained as part of a new Q study should be re-ranked using a similar method to the Q-block approach (in this case simply ranking the whole array from 1 to 16 or 1 to 31 depending on which set is being used). As will be seen below, very good results can be obtained simply using the subset of Q responses.

Comparing the Approaches

Figure 2 graphs the results of the four approaches to deriving factor loadings from Dryzek's original study based on the block method and the *z*-score method, with each approach using different statement sample sizes. The figure shows the average factor loadings from the original study in the black bars. The block method results are shown in the solid grey bar and the two *z*-score method approaches, using the 16 block statements and all 31 shared statements, are shown by the horizontal and angles hatched bars respectively.

The results are clearly different between the methods. The Q-block method produces either very strong average loadings (Factors B and C, which is strong and negative) or very weak (A and D). The *z*-score method produces results that are generally closer to the original study, except for considerably higher average factor loadings for D (using both combinations of statements). All methods suggest that factor B (inclusive republicanism) remains relatively strong among the CP participants, but the results for the remaining three factors are very different.

Assessment of Methods Using the Original Study Data

So, which approach is the most reliable? And why are the results so different using the Q-block and *z*-score method? To answer the first

question, the same methods used above to benchmark the CP data are applied to the original Dryzek data to see which approach most closely replicates the factor loadings obtained by the original analysis. The resulting derived factor loadings were then correlated to the original factor loadings to obtain a measure of accurcy. The results are shown in Table 4. The table shows that, on average, the Q-block method using all 10 blocks produces derived factor loadings that are correlated to the original loadings with a (Pearson's) r of 0.64. This is lower than the reliability cited by Talbott (1963) of 0.7.

By contrast, the z-score approach produces higher correlations to the

Figure 2. Comparison of Original Study Average Factor Loadings to CP Study Derived Factor Loadings (All Four Approaches)

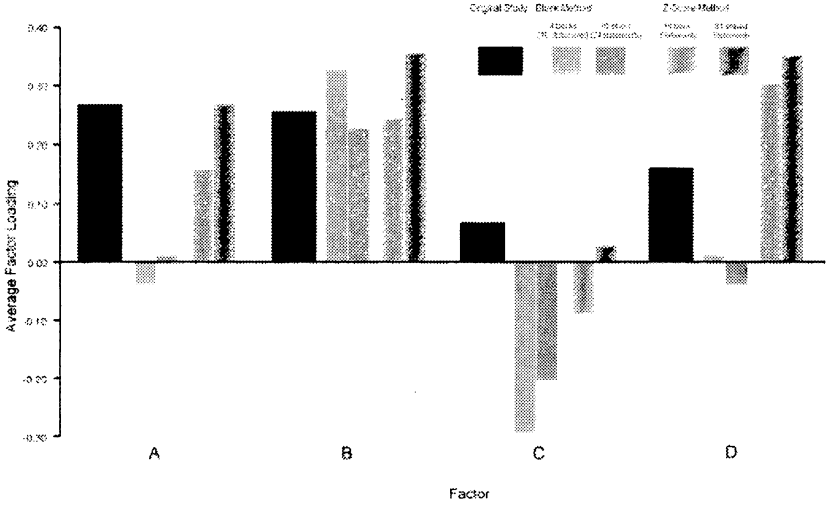


Table 4. Comparison of Concordance to Original Study Factor Loadings with the Different Methods for Deriving Factor Loadings

	A	*B*	*C*	*D*	*Average*	*% of individuals correctly associated*
	Block Method					
Four Blocks	0.76	0.62	0.71	0.34	***0.61***	***68%***
Ten Blocks	0.81	0.70	0.71	0.35	***0.64***	***75%***
	Z Score Method					
16 Statements	0.82	0.72	0.89	0.59	***0.76***	***77%***
All Shared Statements	0.92	0.88	0.94	0.68	***0.85***	***80%***

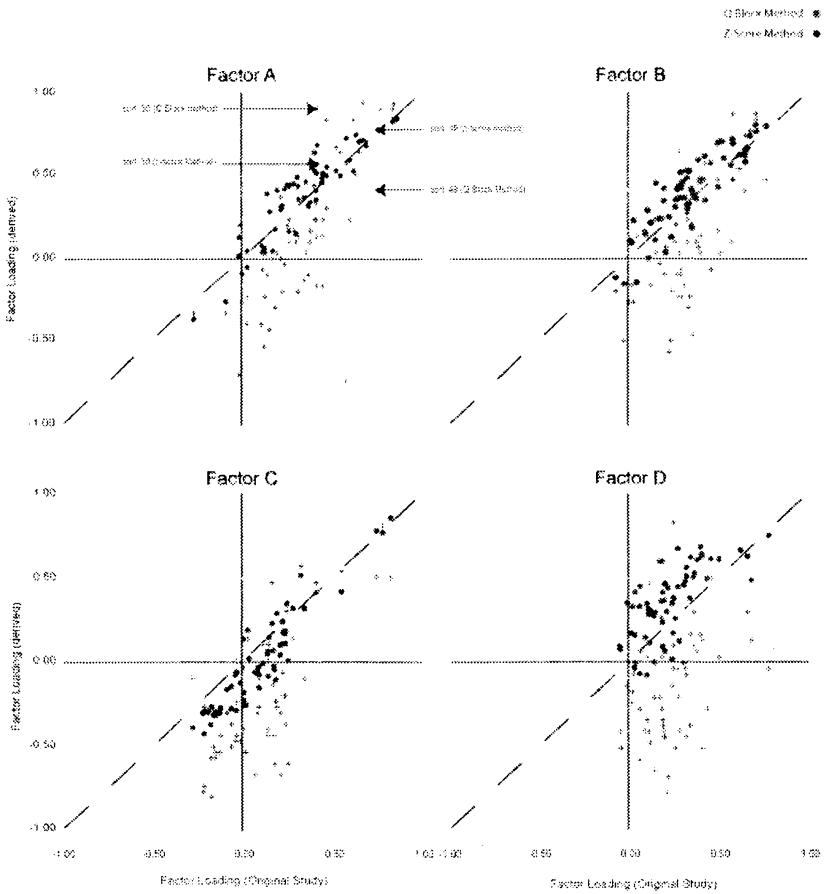
original factor loadings for both the small subset of 16 statements used in the short set of four Q blocks (reliability of 0.76) and the whole set of shared statements (0.85), with the latter being the most reliable of all the methods. The level of reliability is similarly high for the *z*-score method when associating individuals to factors, as is shown in the rightmost column in Table 4.

I will now take a closer look at the results of the two methods, focussing on the 10 Q-block example for the block method and all 31 shared statements for the *z*-score method. The approach involves comparing the level of variation using each method to the original factor loading from the Dryzek study for each individual to see whether there is anything systematic in the differences. The actual method involves plotting the factor loading from the original study against the derived loading using a scatter plot, where each point represents an individual Q sort. If the proxy factor loading and the original study factor loading for an individual is the same, the points will fall along a 45 degree (1:1) reference line and where the proxy factor loading method produces an overestimate, the plot will tend to fall above the 45 degree line, and vice versa for underestimates compared to the original study.

Figure 3 plots the factor loadings for the 60 participants from the original Dryzek study against the derived factor loadings using the two methods for all four factors—with the Q-block results represented as crosses and the *z*-score method represented as dots. It is clear from the figure that the *z*-score method provides the closest approximation to the original factor loadings (this is already known from Table 4). What is interesting is the tendency for the Q-block method loadings to consistently underestimate factor loadings where the original loadings are relatively low. So there does appear to be something systematic going on in relation to the error associated with the proxy factor loadings using the Q-Block method.

It is possible to take a closer look at what might be causing this systematic effect by focussing on two examples from Factor A where the difference between the factor loading estimate is the greatest. Sort number 49 is one of the most strongly loaded on Factor A in the original study. The *z*-score method achieves a very close approximation to the original factor loading, but the Q-block method underestimates the factor loading to the extent that that sort would no longer be associated with the factor at the level set by Dryzek (0.45). The reason for the difference can be explored by plotting the responses in sort 49 against the *z*-scores for Factor A and identifying the relationships between the statements used in the Q block. This is done in Figure 4. The figure plots the z-score (x-axis) from the original Dryzek study for factor A for each of the statements against the (forced) response obtained from Q sort 49 from the original study.

Figure 3: Q-Block Method vs Z-Score Method



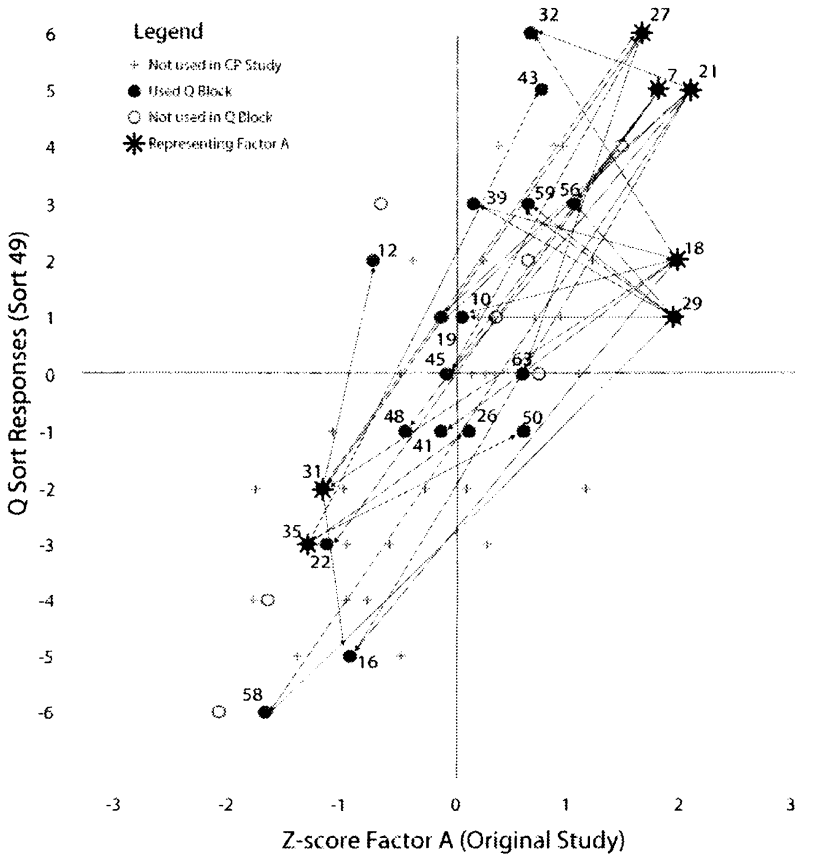
Explaining the Difference in Results

Since sort 49 produced a relatively high factor loading against factor A, it is not surprising that the figure shows a neat linear relationship between the responses for sort 49 and the *z*-scores for Factor A. A similar relationship remains when those statements (represented by the small crosses) that are not common to the original Dryzek and CP study are omitted, which is why the *z*-score method produces a similar result to original factor loading..

However, a different outcome emerges when looking at the same data through the lens of the Q-block method. It can be recalled that, when estimating association with a factor, the Q-block method involves ranking statements associated with a particular factor within a block against the remaining statements in that block. So the method is very

sensitive to the particular orderings within a block, whereas the actual factor loading (and the z-score method) takes into account the relationship between statement responses across all the available statements.

Figure 4: Plot of Sort 49 Responses versus Factor-A Z-scores



The effect of this limitation of the Q-block approach is visually represented in Figure 4 by linking those (starred) statements representing factor A within each block to the remaining three block statements. The statements representing Factor A fall to the extreme left and right of the figure, depending on whether they belong to 'negative' or 'positive' blocks. The slope of the lines indicates the relationship between statements representing Factor A and the others in that block.

In general, an upward-sloping line (from left to right) will increase

the derived factor loading, because that will mean that the positive block statement representing A is higher ranked for positive statements or lower for negative ones. The more upward-sloping the lines, the higher the proxy factor loadings. In most cases the lines are upward-sloping, except in eight cases (out of 30). In all but one of these cases the relatively low Q-block rank is due to one of two statements (18 and 29) that score relatively low in sort 49.

As a result, sort 49 produces a positive, but relatively weak, association with Factor A compared to the result obtained using both the original factor loading and the *z*-score method. This is all because that sort happens to produce a *relatively* low response to two of the statements representing factor A in the Q block. The emphasis is on the term relative, because the response is still positive for these two statements (18 and 29), both of which are members of positive blocks. In other words, the Q-block method in this case appears to be too sensitive to relative ordering within each block, without taking into account the relative ordering of all the statements used in the indexing study.

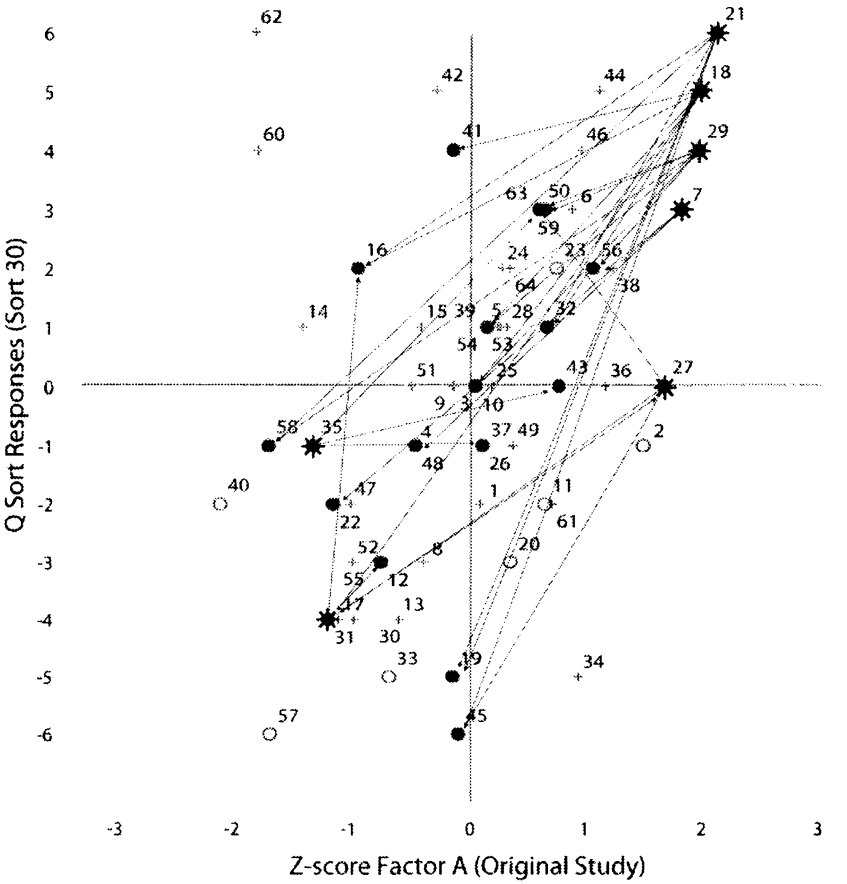
Sort 30 produces a different kind of error in the estimate of the proxy factor loading for Factor A, but for similar reasons. Most cases produce an underestimate (as in the case of sort 49 above). Here we have an overestimate using the Q-block method. The reason for this can be demonstrated in reference to Figure 5 where responses in sort 30 are plotted against the *z*-scores for Factor A in the same manner as was done in Figure 4. It can be seen that the overall association between *z*-scores for all statements in the original study and sort 30 is not particularly strong, but when statements that are not included in the CP study are omitted (in particular statements 62 and 60) the relationship becomes somewhat stronger, hence the slightly higher factor loading derived using the *z*-score method.

When looking at the relationships within the Q blocks, it can be seen that four of the statements representing Factor A in those blocks score relatively high, two relatively low (in the case of the negative blocks) and only one (statement 27) with a mixed relationship, producing a single downward-sloping line out of all 30, hence the very strongly positive proxy factor loading using the Q block method.

Thus, in the example of sort 49 we have a Q sort that is strongly associated with Factor A in the original study, as does the *z*-score method. But the block method underestimates the association because two of the statements representing that factor in their respective blocks happen to have a relatively low score compared to the other statements in those blocks. In the case of sort 30 there is not a strong association in the original study, but a strong association using *z*-score and a very

strong association using Q block because of a combined tendency to respond strongly to the statements representing Factor A, at least compared to the other statements in the same block. Q sort 30 is one of the relatively rare exceptions to the rule. Most proxy factor loadings using the Q-block method produced an underestimate, which is increasingly the case as the actual factor loadings decrease.

Figure 5. Plot of Sort 30 Responses versus Factor-A Z-scores



Discussion

From the assessment of the different methods for benchmarking the results of the CP study to Dryzek's original study it appears that the *z*-score method using the full range of 31 overlapping statements is consistently the most accurate. Thus, returning to the analysis of the CP data, it appears that relative abundance of each factor (in terms of factor loadings) has not changed greatly, except for perhaps the final discourse (D: Anxious Egalitarianism)—although this factor is associated with the

lowest level of confidence. It is very interesting that factors A, C and D produce dramatically different results between the block method and the *z*-score method for the CP data—a far greater difference than is indicated by the relative accuracy of the approaches when the original data is used. This could reflect the fact that the actual composition of factors has changed, thus exacerbating the tendency for the Q-block method to underestimate associations with factors because of the hypersensitivity of the method to the relative location of the statements within a given block. It seems be the case that CP participants who ended up being associated with those factors (according to the *z*-score method) agreed with the overall composition of the factor, but, as we have seen, the relative location of the representative statements can easily change if they are slightly repositioned.

If we are to accept factor loadings as the method for benchmarking individuals to factors, the *z*-score method is again the more accurate. It appears that the power of the Q-block method is diminished by the fact that comparisons are only made within each block and not all the statements that are used in the benchmarking study—in the case of the CP study comprising 31 statements out of the original 64. As such the block method appears to be subjected to stronger random effects. Using all 10 blocks the Q-block method associated individuals with factors the same way as for the original study in 75 percent of cases, whereas the *z*-score method achieved 80 percent.

However, the differences in results raise an interesting issue regarding the status of factor loadings as the primary measure for indexing the association of individual Q sorts with factors. Whereas the *z*-score method appears to better approximate the factor loadings as they would have been calculated in the original study, by focussing on statements that actually distinguish a factor from the others, the Q-block method is not necessarily inferior in terms of determining levels of association with a factor, because it tends to omit statements with low *z*-scores as part of the measure of association.

The relative merit of each approach is ultimately resolved by the way in which association with a factor is defined (or the ontological status of an association with a factor). Q methodology uses a threshold factor loading to associate individuals with factors, and here the *z*-score method is clearly the more accurate proxy method.

It should be noted that the block method would probably produce better results as a proxy for a factor loading in studies that involve factors that involve a high proportion of unique *z*-scores, which was clearly not the case for the original Dryzek study. The *z*-score method is far less sensitive to this problem.

One disadvantage in using the *z*-score method for indexing the

association between individuals and factors could be the relative difficulty in implementing the necessary survey (which in this case involved a complete Q sort) and obtaining results for use in a *z*-score correlation. The block method as suggested by Talbott (1963) is less demanding on participants insofar as it only involves a series of rankings within groups of four statements (where there are four factors being benchmarked). The CP study involved obtaining scores as part of a wider Q study. It could be possible to simply obtain unforced data using Likert responses, but doing so changes the conditions of instruction compared with the original study, unless it also involves the analysis of unforced data. Another approach could be to simply ask the respondent to rank the statements in order of agreement. This is something that could be feasible with relatively small sets (of 16 statements or less), but as the analysis above shows, the smaller the sample, the lower the confidence in the findings. Still another approach could be to use the block-ranking method suggested by Talbott, but with the addition of an unforced Likert response, that could be used to weight the responses to produce a distribution.

In any case, the analysis above has shown that in cases where a new Q study overlaps with an original study and the researchers seek to compare results while still retaining the ability to undertake a new Q analysis, the *z*-score method using data from overlapping statements between studies appears to produce very high quality results compared to the Q-block method.

Conclusion

Talbott (1963) has done a great service by advancing the possibility suggested by Stephenson that the factors obtained using Q methodology could be benchmarked to a wider population. This paper has explored the use of the method that Talbott proposed in the context of a Q study that has a shared set of statements with a previous study, in this case Dryzek's (1994) study into Australian discourses of democracy. It has compared the accuracy of the block method to a simple correlation between responses to shared statements and the corresponding *z*-score from each factor in the original study (the *z*-score method). The results suggest that the latter method is not only conceptually and analytically simpler, but also considerably more accurate—at least in cases where there are relatively few unique statements from the original study based on *z*-scores. However, while the accuracy claim is straightforward in relation to factor loadings, when it comes to associating individuals with factors the situation is less clear-cut.

The simplicity of implementing the Q-block method via a series of ranking exercises does make it more appealing for adoption using large numbers. Yet, there are methods that could be similarly straightforward

that also produce data amenable for the *z*-score method.

This is clearly an area where more work needs to be done, including a more formal investigation into why the methods used here produce markedly different results. This paper has attempted to advance the development of a benchmarking methodology for Q—an important field of study where there is still considerable work to be done.

References

Dryzek, J. S. (1994). Australian discourses of democracy. *Australian Journal of Political Science, 29,* 221–239.

Hackert, C. & Braehle, G. (2006). FlashQ (computer software). Available from http://www.hackert.biz/flashq/home/.

Talbott, A. D. (1963). The Q-block method of indexing Q typologies. Presented at the AEJ Conference, Lincoln, Nebraska. [reprinted in this issue, pp. 6–24].

Appendix: Factor Scores: Australian Discourses of Democracy

No.	*Statement*	*A*	*B*	*C*	*D*
1	Democracy for me means not so much freedom, but more flexibility; to adapt to situations.	0	−1	−1	0
2	Democracy is a term which has lost much of its original meaning.	4	−2	−3	−3
3	There are three stages to a revolution: protest, participation in the democratic process, and fulfilment of aims.	−1	−2	−5	0
4	The concept of democracy is the greatest good for the greatest number.	−2	0	1	2
5	We can certainly take advice from the experts. But in the end it's one of the rights in a democracy that each citizen has his capacity to influence.	0	3	3	−1
6	Real national leadership does not lie in exploiting the unpopularity of a failed government, but in addressing national problems in a realistic way and in defining a strategy to overcome them.	3	3	5	4
7	If you want your vote to count for anything, then you need to not vote in ignorance, you need to know something about what you're being offered.	5	5	2	5
8	The more emphasis that is placed on the rights of minorities, and the need for affirmative action to enhance those rights, the more is the concept of democracy and the rights of the majority in danger of being weakened.	−1	−4	2	−2
9	Idealism is the spirit of opposition.	−1	−2	−5	−1

No.	*Statement*	*A*	*B*	*C*	*D*
10	The true believers are about decency in life, about a capacity to debate the issues and to receive different views.	0	3	5	3
11	We as individuals have a right, that's what democracy means to me	2	−1	0	−2
12	Republicanism is a smokescreen for democratic treachery, incompetence and lunacy.	−2	−6	2	−4
13	Fairness has had to be redefined from equality of opportunity and equal application of laws to equality of impact on socially unequal groups.	−2	2	−1	0
14	An aspect of democracy is we can hold our representatives accountable.	−4	2	1	1
15	Surely equality means equality, none are more equal than others. We are all Australians, aren't we?	−1	−1	0	3
16	Until women are equally represented in Australian parliaments, we will not have a genuine representative democracy.	−3	6	−3	−6
17	Constitutional monarchy has given us protections, freedoms, and rights that do not exist in republics.	−4	−5	−1	−1
18	There's a dollar democracy that runs through our supposed democracy.	6	1	−2	2
19	We are a democratic country.	−1	−1	−2	6
20	The fundamental dynamic of Australian society is that it is one of evolution, not revolution.	0	0	0	1
21	In Australia the rich have virtually unlimited access to the legal system and the capacity to use it to achieve their own ends.	6	5	−3	−1
22	You don't have to wait three years to say 'I'm not happy with the job you did.' You can apply pressure, and enough people writing to the local member can turn the situation around.	−2	0	4	−3
23	I don't think we have a lot of input into legislation and a lot of things that are important.	2	2	−2	−2
24	The unemployed have become effectively disenfranchised.	1	1	−1	0
25	Governments like to deal with big blocks of numbers, so they can do deals and fix things up without having to talk to a hundred thousand people individually.	0	0	1	1
26	Our government is attempting to mould our society to the needs of a profit-oriented market	0	0	−4	3
27	I'm always cynical about government processes.	5	1	−4	0
28	I don't think the freedom of a right to be poor, and underrepresented, is a kind of freedom this country wants to promote.	1	2	1	−4

No.	Statement	A	B	C	D
30	The Australian way is now established as one of a cooperative effort between the state and organised labor.	-3	-2	-1	0
31	The only time we are all equal is when we vote in the ballot box.	-4	1	-2	4
32	A lot of people may be politically ignorant, but they believe what they see and hear on television, and that makes them vote the way they do. They have been brainwashed by the clever advertising people.	2	6	3	1
33	We live in a great democracy.	-3	-3	-4	-2
34	Our form of democracy is not perfect, but it beats by a long shot any other alternative we've been shown.	2	-2	2	4
35	Our political system's strength is that it can change to meet changing circumstances.	-4	-3	1	-1
36	All parties seem to be appalling.	4	-2	-4	-4
37	This is one of the great things about living in a democracy. You have the right to express an opinion even if you are wrong.	0	1	4	2
38	The problem with democracy in Australia is that we are rarely asked our opinions.	4	2	-1	-2
39	Sometimes I get a bit worried that there's a lot of small minority groups that tend to get a lot of the say	0	-1	4	6
40	I don't know too much about democracy. I feel ridiculous voting.	-6	-5	-6	-6
41	Governments can talk a great game about 'democracy'. They fall down, in practice, because they cannot accept any sharing of the power to influence decisions with those who might have different motives and ideas.	1	1	6	-1
42	The whole thrust of contemporary politics is so preoccupied with the pursuit of economic rationality that its proponents have lost any sense of a larger frame of value or ethical orientation that might properly guide political and social life.	-1	4	1	0
43	I don't think there's anything wrong with people voting for their self-interest.	3	-4	-1	3
44	The simple-minded instinct of the layman may be a better guide to policy than the clever ideas of some over-subtle economist.	3	-1	3	2
45	Free markets work because individual people, cooperating peacefully and voluntarily through markets, can achieve much that politicians and bureaucrats cannot achieve using compulsion and direction.	-1	-3	5	1

No.	*Statement*	*A*	*B*	*C*	*D*
46	If the majority is an oppressive majority it can't be fair, it can't be a democracy	3	4	-3	-1
47	In societies that are functioning well, old constitutions tend to be good constitutions, because people have learned to adjust their affairs to them.	-3	-4	1	-3
48	Low feminist participation rates compound the domination of the public realm by males.	-2	3	0	-2
49	I'd like to see real democracy stand up and say 'here I am!'	1	-1	0	-1
50	The last thing we need is a Canberra 'Big Brother' telling us what we should read and hear.	3	1	1	-3
51	Instead of having representative democracy we could have proportional representation so if a party got 5% of the votes it could get 5% of the seats.	-1	0	-5	2
52	My preference would be for a bureaucratic form of government, where even organised crime couldn't afford to buy up the whole public service.	-3	-3	0	1
53	We need to be in every sense, including the symbolic one, our own masters.	1	0	-1	-3
54	Freedom is fine, but you don't have a right to rights	1	-1	3	1
55	The people who are experts in a field may have a duty to educate people	-2	-5	-2	-4
56	I don't want to be in a place where only a minority's viewpoint gets the right to say what's going to happen.	4	0	2	5
57	The true function of government is to maintain peace and justice. This does not include interfering in national or international trade or commerce, or in the private transactions of its electors save only as they threaten peace and justice.	-5	-6	3	-5
58	If Australia is to recover from the horrible mess we are in at present, we will need leaders of high principles, who will govern with honesty, justice, and righteousness, fearlessly putting into practice the principles of the Bible.	-5	-3	6	-5
59	Our community and nation should invoke a spirit that you are important, that you do have a way of influencing things.	2	5	2	2
60	The working class has one historic duty: to overthrow capitalism and establish a workers' republic under the red flag of freedom.	-5	-4	-2	-5

No.	*Statement*	*A*	*B*	*C*	*D*
61	Our society will be more secure and more attractive if it is also more equitable and humane, as well as more productive and efficient.	1	4	0	5
62	Until we white Australians give back to black Australians their nationhood, we can never claim our own.	−6	1	−6	1
63	Concluding that some races/cultures are better than others must be avoided.	2	4	0	3
64	People have to make sure that everyone has a say. Some people are bullies, and they'll shout and will be heard above everybody else. The shy ones shouldn't be downtrodden.	1	2	4	0