A Robust Transformation Procedure for Interpreting Political Texts*

Lanny W. Martin
Assistant Professor
Department of Political Science
Rice University
PO Box 1892, MS 24
Houston TX 77251-1892
lmartin@rice.edu

Georg Vanberg
Associate Professor
Department of Political Science
University of North Carolina, Chapel Hill
Chapel Hill, NC 27599-3265
gvanberg@unc.edu

Abstract

In a recent article in the *American Political Science Review*, Laver, Benoit, and Garry propose a new method for conducting content analysis. Their *Wordscores* approach, by automating text coding procedures, represents a fundamental advance in content analysis and will potentially have a large long-term impact on research across the discipline. In this research note, we contend that the usefulness of this procedure is unfortunately limited by the fact that the transformation procedure used by the authors (which is meant to allow for the substantive interpretation of results) has two significant shortcomings. Specifically, it distorts the metric on which content scores are placed—hindering the ability of scholars to make meaningful comparisons across texts—and it is very sensitive to the texts that are scored—opening up the possibility that researchers may generate, inadvertently or not, results that depend on the texts they choose to include in their analyses. We propose (and have written program code to implement) a transformation procedure that solves these problems.

* We would like to thank Ken Benoit for extensive comments and suggestions on a previous draft.

In a recent article in the *American Political Science Review*, Laver, Benoit, and Garry (2003) [hereafter, LBG] propose a new method for conducting content analysis.[1] The thrust of their proposal is to estimate the ideological position expressed in a text by treating the individual words in that text as "data" to be scored rather than as words to be understood. By automating the scoring procedure instead of relying on individual coders to make judgments about the meaning of a text, the LBG approach offers three significant advantages over traditional approaches: (1) It can be quickly and cheaply implemented on a conventional computer; (2) it eliminates the problems created by divergent coder judgments; and (3) it dispenses with the need to understand the language in which a text is written. This approach represents a fundamental advance in how to conduct content analysis, and it would be difficult to overestimate its potential long-term impact on research across the discipline. We agree with the authors that "there is no reason the technique should not be used to score texts generated by participants in any policy debate of interest, whether these are bureaucratic policy documents, the transcripts of speeches, court opinions, or international treaties and agreements" (326). Already, researchers are beginning to employ this procedure to analyze the judicial process in the US (McIntosh, Evans, and Cates 2005; McGuire, Vanberg, and Yanus 2006), presidential campaigns in France (Laver, Benoit, and Sauger 2005), intra-party politics in Italy (Giannetti and Laver 2004), and constitution writing in the European Union (Benoit *et al.* 2005).

In its current form, however, this approach has a significant limitation that will prevent scholars from fully exploiting the advance offered by LBG. The Wordscores procedure generates two important ideological measures for a text: a "raw score" and a "transformed score." The transformed score is central, because it allows substantive interpretation of the results generated by the procedure. Unfortunately, the particular transformation offered by LBG suffers from two weaknesses. First, the transformed scores are not robust to the set of texts that are scored. As we shall discuss, this may be problematic even under the best of real-world conditions, but it will be especially so when an analyst is

[1] The authors have made their "Wordscores" procedure (implemented as a suite of programs written for STATA 7.0 and higher) available online at http://www.wordscores.com/.

forced to choose a subset of potentially relevant texts. This is likely to be the case for the analysis of legislative speeches, for example, as well as for most of the other policy-related texts referenced above. This particular feature of the LBG transformation opens up the danger that researchers may generate, inadvertently or not, results that depend on the set of virgin texts they choose to include in their analysis. Second, the LBG transformation presently does not allow researchers to make direct comparisons across all of the texts in the analysis. This implies that we cannot currently use this approach to investigate such questions as whether the stated policy stances of parties are changing over time or whether political actors are as divided on one issue as they are on another. In this note, we propose (and implement) an alternative transformation procedure that resolves both these problems and that will hopefully make the advance offered by LBG even more useful to scholars.[2]

THE WORDSCORES PROCEDURE

LBG's fundamental insight is elegant. The essence of the approach is to use texts with an exogenously defined ideological position (the "reference" texts) to estimate the ideological position of texts whose position is unknown (the "virgin" texts). To do so, the LBG procedure creates a "dictionary" that assigns each word that appears in the reference texts an ideological score. This score is an average of the ideological scores assigned to the reference texts, weighted by the relative frequency of the word across the reference texts. With this dictionary in hand, the ideological position of any virgin text is estimated as the frequency-weighted average position of the dictionary words appearing in the virgin text.

Before discussing the drawbacks of certain features of the LBG approach, and our proposed modifications to them, it is useful to outline the Wordscores procedure in more detail. While our presentation diverges slightly in motivation (and notation) from the LBG treatment, the creation of "raw scores" is computationally equivalent to theirs. Suppose we have *n* reference texts whose ideological position on some unidimensional scale is known. Let A_r denote the ideological score of reference text *r*

[2] As we discuss below, we have also written a STATA program that implements the new transformation, which can be used in conjunction with LBG's Wordscores program to perform textual analysis.

on this dimension, where $r \in \{1,2,...n\}$. (For example, if text T1 expresses a position to the left of T2, we might assign T1 a score of 0 and T2 a score of 1.) Let N_{wr} denote the number of times that the word *w* appears in text *r*. The relative frequency of each word *w* across the reference texts, R_{wr}, is given by:

$$R_{wr} = \frac{N_{wr}}{\sum_{j=1}^{n} N_{wj}}$$

Taken on its own, R_{wr} tells us the frequency with which word *w* appears in reference text *r* relative to the other reference texts. Suppose there are two reference texts and we pick a word out of one of the texts at random. Then—once we have learned what the word is—we try to guess which of the two texts the word came from. Loosely speaking, R_{wr} provides our best guess: It is the probability that a given word *w* comes from text *r*, given a random draw and no additional information.[3]

Since the reference texts typically will be of different lengths, these relative frequencies must be standardized by the length of each text before calculating ideological scores for individual words. Failure to do so would imply that simply lengthening a reference text would have an effect on the scoring of documents by increasing the relative weight given to the scores of longer reference texts. This standardization is achieved by dividing the relative frequency of word *w* for text *r* by the total number of words in text *r*, denoted by N_r:

$$R_{wr}^{\text{standardized}} = \frac{R_{wr}}{N_r}$$

From these standardized relative frequencies, we can calculate the weight each reference text score should receive in scoring the ideological position of word *w*:

[3] LBG (2003, 316) label their quantity P_{wr} as providing this probability. This is only accurate if the reference texts are of exactly equal length.

$$W_{wr} = \frac{R_{wr}^{\text{standardized}}}{\sum_{j=1}^{n} R_{wj}^{\text{standardized}}}$$

Although we have arrived at this weight from a slightly different route, it is computationally equivalent to the weight P_{wr} calculated in the first step of the LBG procedure (see LBG 2003, 316). Using these weights, we can create a "dictionary" of estimated ideological positions for the words contained in the reference texts. The estimated score of word *w* is simply the weighted average of the ideological scores assigned to the reference texts. The ideological score, S_w, of word *w* is thus given by:

$$S_w = \sum_{j=1}^{n} W_{wj} A_j$$

Finally, this dictionary of word scores can be used to estimate the ideological position of any text that contains words out of the dictionary. Suppose we want to score text *t*. The procedure is simple: The ideological position of the text as a whole is estimated as the average ideological position of the scored words that are included in the text. Let *S* be the set of scored words that we have obtained from the reference texts and let *T* be the set of words in text *t*. Then the ideological position, P_t, of text *t* is estimated by the following "raw score:"

$$P_t = \frac{\sum_{w \in S \cap T} N_{wt} \cdot S_w}{\sum_{w \in S \cap T} N_{wt}}$$

The raw score assigned by the LBG procedure to any text is thus uniquely determined by the exogenous scores assigned to the reference texts and the relative frequencies of words across the reference texts. Note in particular (this will turn out to be crucial) that we can apply this scoring procedure to the original reference texts used to create the dictionary. Of course, the raw scores for the reference texts will not be equal to the original reference scores that we assigned to these texts. Since raw scores are a weighted average of words contained across the reference texts, they will be dispersed on a much smaller scale than the reference text scores.

AN ASIDE: OVERLAPPING WORDS

At this point, it is useful to take a short detour to address an important implicit assumption in the Wordscores procedure. Simply because of the nature of language, many words are likely to appear in *all* of the reference (and virgin) texts. Because they are common to all reference texts, these overlapping words receive ideological scores that lie between the reference text scores. The more evenly a word is distributed across the reference texts, the more "centrist" the ideological score that Wordscores assigns to the word. One consequence of the centrist position assigned to overlapping words is that raw scores are usually pulled towards the interior of the interval defined by the reference texts scores. Indeed, as we discuss in the next section, it is precisely this "bunching" of raw scores that generates the need for transformed scores that aid substantive interpretation of the results.

By scoring overlapping words alongside *discriminating* words that are unique to specific reference texts, the original Wordscores procedure makes the implicit assumption that overlapping words express a centrist ideological position. In many cases, this is plausible. Parties from across the spectrum may use similar language in espousing centrist positions designed to appeal to moderate voters. However, words may also be shared across reference texts not because they express a centrist position but because they do not carry ideological content (e.g., words like "and," "the," etc.). In this second case, Wordscores treats words that carry no ideological content as expressing a moderate or centrist position. Unfortunately, there is no immediate or easy fix for this issue. It is precisely the fact that researchers do not need to concern themselves with coding individual words or phrases, (or even understanding them) that makes Wordscores such a powerful tool. To separate overlapping words into those that carry and those that do not carry ideological content is an impractical demand that undermines the very purpose of the enterprise.

However, it is possible to perform at least a simple robustness check to investigate the impact of treating overlapping words as either expressing a centrist position or as carrying no ideological content. The danger that overlapping words are simply non-ideological probably increases with the degree of overlap. For example, we would expect words like "the" and "and" to appear in roughly equal proportions across reference texts (once we standardize for length, of course). To gauge the impact of overlapping words on raw scores, we could therefore perform two separate scoring procedures. First, we could score all words in the reference text (the original LBG procedure). Second, we could eliminate words with a high degree of overlap and score only those words that discriminate sufficiently between reference texts. The first estimates provide scores under the assumption that overlapping words express a centrist position; the second provides scores under the assumption that they do not carry ideological information. Our revised Wordscores program implements this option. Researchers can specify a specific degree of overlap such that only words that do not exceed this overlap cutoff are used in the scoring procedure.

INTERPRETATION OF VIRGIN TEXT SCORES

The substantive interpretation of raw scores is a crucial issue but, as LBG point out, not a straightforward one. Overlapping words pull raw scores towards the interior of the interval defined by the reference texts scores. As a result, raw scores cannot be directly compared to the exogenous scores attached to the reference texts. Moreover, the "bunching" of raw scores places them on an unintuitive metric that makes direct interpretation of the virgin text raw scores difficult. Consequently, raw scores must be transformed in some form to allow substantive interpretation. At the very least, researchers need to calculate ratios of raw scores to express their relative cardinal positioning – itself a transformation.

LBG propose a transformation that provides raw scores with the same dispersion as the reference text scores by centering the raw scores around their mean and adjusting the spread of the scores to correspond to the spread of the reference texts. The LBG transformed score, P_t^* , of text *t* is given by:

$$P_t^* = (P_t - \overline{P}_v)(\frac{SD_r}{SD_v}) + \overline{P}_v$$

where $\overline{P}_v$ is the average raw score of the virgin texts and SD_r and SD_v are the standard deviations of the reference and virgin text scores, respectively.

As LBG point out, theirs is only one possible transformation procedure (2003, 316f.). In what follows, we propose an alternative. To understand the advantages of our proposal, it is necessary to discuss two (related) drawbacks of the particular transformation chosen by LBG. The first is that *the* *transformed scores assigned to virgin texts depend on the particular combination of virgin texts that are scored*. We view this as potentially problematic for any research design, but we believe it will become especially dangerous as scholars move beyond the content analysis of manifestos—which comprise a relatively small number of texts that are sometimes (though not always) available for all parties competing in a given election—to the analysis of legislative speeches, newspaper reports, and so on, which are sufficiently abundant in number that researchers will almost certainly have to make choices about which subset of them to include in their analyses. The second drawback is that *the transformation procedure cannot recover the original scores assigned to the reference texts*. This implies that we cannot make meaningful comparisons across reference texts and virgin texts, obviously a fundamental limitation even for scholars lucky enough to have a complete set of virgin texts.

1) The transformed scores depend on the combination of virgin texts scored

Consider the following example, which extends an example used by LBG in their original article. The example uses British party manifestos from the 1992 general election, along with exogenous scores (on a 20-point scale) on an "economics" dimension for these manifestos derived from an expert survey (Laver and Hunt 1992), to estimate the positions of British party manifestos during the 1997 general election. Following LBG, we assign the Labour manifesto in 1992 a reference score of 5.35, the Liberal Democrat manifesto a score of 8.21, and the Conservative manifesto a score of 17.21. Using these texts to score the three manifestos for the 1997 election, the LBG procedure yields transformed scores of 9.11 for Labour, 5.00 for the Liberal Democrats, and 17.17 for the Conservatives, scores that sensibly reflect the relative changes in party positions over this time frame (LBG 2003, 320).

Suppose that instead of scoring all three 1997 manifestos, one were to vary the set of virgin texts. For example, assume that a researcher is interested in figuring out whether (or how) the main two parties competing on the center-left—Labour and the Liberal Democrats—ideologically "repositioned" themselves *relative to one another* following the narrow electoral defeat suffered by Labour in 1992. Given her question (and certain assumptions about how parties compete), this researcher might feel quite comfortable excluding the 1997 Conservative party manifesto from her analysis. Unfortunately, if she were to do so using the current LBG transformation, she would produce radically different ideological positions than those obtained above. Specifically, the LBG transformation in this scenario would assign a score of 5.91 to the Liberal Democratic manifesto and a score of 14.66 to the Labour party. In other words, Labour would now look as though it has moved quite far to the right—in fact, much closer to the LBG estimated position for the 1997 Conservatives than to the Liberal Democrats. Clearly, this paints a dramatically different picture of British politics for this period than most experts would accept.[4] Other combinations of virgin texts would yield still other scores.

It is easy to see why transformed scores under the LBG procedure depend so heavily on the specific set of texts that is scored. To adjust the dispersion of the raw scores, their transformation relies on the standard deviation of the virgin text raw scores. The standard deviation, of course, depends on the particular set of virgin texts that are scored. Put simply, the LBG transformed scores are not robust to the selection of virgin texts. In some circumstances, of course, the set of virgin texts will be defined in a natural way, e.g., all of the parties running in an election campaign.[5] But in many applications, it will *not*

---

[4] We also note these new scores for Liberal Democrats and Labour fall outside the upper confidence bound of their scores when all three virgin texts were used (very far outside for Labour).

[5] Importantly, though, it is very rarely the case that scholars using election manifestos will find themselves in such happy circumstances. Continuing with the British case, for example, more than forty other parties besides the three discussed in the text competed in the 1997 elections (ten of them winning seats). Unfortunately, most of these parties do not have easily retrievable manifestos. The availability of texts will generally be even worse if we wish to extend our analysis back in time or to more fragmented party systems. Even the largest cross-national collection of election manifestos (from the Comparative Manifestos Project [CMP]) has numerous gaps, some unavoidable and some brought about by the decision of CMP researchers not to collect manifestos of parties deemed (retrospectively) as “one-hit

be obvious what the appropriate set of virgin texts to score is. For example, for someone interested in analyzing the content of parliamentary speeches, it may not be apparent which speeches should be included in the analysis—all speeches? just those of party leaders? should speeches by independents be included? should a short interjection during general debate count as a separate speech? As another example, consider applications in judicial politics. Given the overwhelming number of documents, and the fungible lines between legal areas, scholars will have to select a subset of judicial opinions or briefs to analyze. Because of the sensitivity of scores to the set of texts, the *choice* of which texts to include or exclude could—consciously or unconsciously—have significant effects on the position attached to any particular text. To increase confidence in an analysis of transformed scores, it would therefore be desirable to develop a transformation that makes scores *independent* of the particular set of texts scored.

2) The transformed scores of the reference texts do not correspond to the original scores of the reference texts

As LBG point out, a primary reason for transforming raw scores is to make comparison of virgin text scores with the original reference text scores possible. To enable such direct comparison, it is crucial that any transformation generates transformed scores for the *reference* texts that correspond to the original scores assigned to them. That is, the transformation procedure must be able to "recover" the original reference text scores. If it cannot, then the transformed scores are *prima facie* on a different metric than the reference text scores. Unfortunately, the LBG procedure is unsuccessful in this regard.[6] Continuing with the British example, recall that LBG used three 1992 reference texts, which were assigned a priori scores on the economics dimension based on the Laver-Hunt expert survey, to create the ideological

---

wonders" or "non-coalitionable." Our suspicion is that such systematic exclusion of manifesto texts by CMP will only serve to worsen the problems we have identified with the use of the LBG transformation.

[6] This is true if the raw scores of the reference texts are included alongside the raw scores of the virgin texts to calculate the standard deviation and mean used for the transformation, or if only the virgin texts are included, with the reference texts scored separately.

“dictionary” used for coding the 1997 virgin texts. As described earlier, this automated coding procedure generates the raw scores for the 1997 texts, which are then transformed by the LBG procedure. If we were to use to this same transformation to score the original 1992 *reference* texts, then, at a minimum, we should expect to recover the a priori scores from the Laver-Hunt survey (5.35 for Labour, 8.21 for Liberal Democrats, and 17.21 for Conservatives). The scores that are actually generated by the LBG procedure, however, are 2.19 for Labour, 6.34 for the Liberal Democrats, and 20.04 for the Conservatives. These transformed 1992 reference text scores obviously do not match with their a priori scores (in fact, the a priori scores lie outside the respective 95% confidence intervals of the transformed scores). The transformed score for the Conservatives is even outside the range of the twenty-point Laver-Hunt scale. Clearly, when we cannot accurately compare the original reference text positions with their own scores, we cannot meaningfully compare them with the scores of the virgin texts.

A ROBUST TRANSFORMATION PROCEDURE

The transformation procedure applied to the raw scores should have two features. First, it should be internally consistent. The transformed scores should recover the scores assigned to the reference texts. Otherwise, it is impossible to compare virgin texts to the reference texts because the scores are, by definition, on a different scale. Second, it should produce scores that do not depend on the set of virgin texts. That is, it should eliminate the sensitivity of scores to the choice of virgin texts. The scores that are produced should be unaffected by the availability of texts or by the (conscious or unconscious) choice of which virgin texts to include in the analysis. We propose a transformation that satisfies both of these criteria. We also supply the appropriate STATA program that can be used in conjunction with LBG’s original Wordscores program to calculate the alternative transformed scores.[7]

The intuition behind our transformation is simple. We start with two reference texts (specifically, with texts that have a priori scores at opposite ends of the dimensional scale) to create the dictionary that is used to generate the raw scores for all texts, including the reference texts. We then “stretch” the

---

[7] The program, along with instructions, is available at http://[omitted].

distance between the raw scores of the reference texts to correspond to the original metric while preserving the *relative* placement of the virgin texts. To do so, we need to "center" the raw scores on the same scale as the assigned reference scores, and then increase the spread of the raw scores to correspond to the original spread of the reference scale. Given raw score P_t, the following transformation yields the proper transformed score $\hat{P}_t$ (recall that A_r is the assigned score of reference text *r*):

$$\hat{P}_t = \{(P_t - P_1)\frac{A_2 - A_1}{P_2 - P_1}\} + A_1 \ .$$

To see what this transformation does, consider the scores it will assign to the two reference texts, R1 and R2. The transformed score assigned to reference text R1 is given by:

$$\hat{P}_1 = \{(P_1 - P_1)\frac{A_2 - A_1}{P_2 - P_1}\} + A_1 = A_1 \, .$$

That is, the transformed raw score is equal to the original reference score assigned to R1. Similarly, the transformed score of reference text R2 is given by:

$$\hat{P}_2 = \{(P_2 - P_1)\frac{A_2 - A_1}{P_2 - P_1}\} + A_1 = A_2 \, .$$

The transformation thus recovers the original scores assigned to the two reference texts. Any virgin text will be placed on the same scale in the proper relative position to these reference texts, where this relative position is derived from the spread of the raw scores.[8]

At this point, it is important to draw attention to one feature of this transformation, *or any transformation*, that recovers the original scores assigned to the reference texts. Such a transformation is always limited to working with *two* reference texts as "inputs." The reason is immediate. To recover the original scores, the procedure needs to (1) stretch the raw scores to the original metric and (2) preserve

[8] Like the LBG procedure, our procedure (and program) generates standard errors and confidence bounds for the raw and transformed virgin text scores that are produced in the manner outlined by LBG (see LBG 2003, 317).

the proper relative distance between scores. In general, achieving (1) and (2) simultaneously is only possible with two reference scores. The reason is that the relative distance between *two* scores imposes no constraint on adjusting the distance between these scores. As soon as three or more scores are involved, however, the fact that the relative distances of raw scores generally do not correspond to the relative distance of reference scores makes a transformation that achieves both (1) and (2) impossible. Consequently, employing our alternative transformation procedure demands restriction to two reference texts in generating the word dictionary.

AN EXAMPLE

To illustrate the difference between our transformation and the LBG transformation, we present a brief example in this section. Consider four texts that are made up of words indicated by the letters A, B, C, D, and E. The composition of the texts is listed in the top panel of Table 1.

<<**Table 1 about here**>>

Suppose we know that reference text R1 expresses a liberal position and R2 expresses a conservative position, and we assign these reference texts scores of 0 and 1, respectively. The bottom panel of Table 1 lists the raw and transformed scores that would be assigned to each text. Recall that our transformation is unaffected by the set of virgin texts that are scored. As the table shows, this transformation recovers the original reference scores and assigns sensible scores to the other texts. T2 and T3, which are composed of an exact balance of words that appear in the two reference texts, are assigned a score of .5, halfway between the reference texts. T1 and T4, which draw primarily on words associated with one of the reference texts and have little material to “balance” their score are assigned scores to the left and the right of the reference texts, respectively. Finally, T5 is almost identical to R2, but makes use of an additional word shared across R1 and R2. Accordingly, the text is placed just to the left of R2.

As we showed above, the LBG transformed scores depend on the particular set of virgin texts that are scored. We therefore report two alternatives the table. The first row shows the transformed scores if all five virgin texts (T1 through T5) are scored. The second lists transformed scores when only four virgin texts are used (T2 throughT5). Comparing these scores to our scores, several things emerge. First, we can see that the LBG scores are sensitive to the set of virgin texts that are scored. The scores of T2, T3, and T5 move substantially as we include or exclude T1. Moreover, the scores also demonstrate the difficulty in comparing LBG scores to the original reference scores. Depending on which virgin texts are scored, the LBG transformation assigns T2 and T3 scores of either .43 or .62. Aside from the movement in these scores, their values indicate immediately that they cannot be readily compared to the reference text scores. If they could be compared in this way, the scores should be placed at *exactly* .5, since these texts represent an exact balance of words used across the two reference texts.

CONCLUSION

The content analysis procedure offered by Laver, Benoit, and Garry (2003) offers a revolutionary new method for conducting content analysis that has implications for all substantive areas of political science. By automating the scoring process through a readily available program, LBG allow researchers to analyze texts quickly and cheaply and to avoid many of the difficulties introduced by subjective coder decisions. As a result, Wordscores has the potential to become a standard tool for researchers in the social sciences. To more fully realize this potential, it is obviously important that our content analysis procedures allow meaningful substantive interpretation of texts. In this note to their *APSR* article, we have outlined an addition to the LBG procedure that we hope will make the advance offered by LBG even more useful. The particular transformation for raw scores that we propose remedies several difficulties in the original transformation procedure. First, it eliminates the sensitivity of transformed scores to the particular set of virgin texts that are scored. This is crucial, since it will not be obvious in many applications which particular texts should be scored. Second, the procedure makes the scores of virgin texts directly comparable to the original reference text scores. Both modifications are critical for scholars interested in drawing conclusions about the relationship between the positions of different texts over time and between political actors.

References

Benoit, Kenneth, Michael Laver, Christine Arnold, Madeleine O. Hosli, and Paul Pennings. 2005. "Measuring National Delegate Positions at the Convention on the Future of Europe Using Computerized Wordscoring."' *European Union Politics* 6(3): 291-313.

Giannetti, Daniela, and Michael Laver. 2004. "Party Factions and Split Roll Call Voting in the Italian DS." Working Paper, University of Bologna.

Laver, Michael, Kenneth Benoit, and John Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97(2): 311-31.

Laver, Michael, Kenneth Benoit, and Nicolas Sauger. 2005. "Policy Competition in the 2002 French Legislative and Presidential Elections." *European Journal of Political Research*, Forthcoming.

McGuire, Kevin, Georg Vanberg, and Ali Yanus. 2006. "Targeting the Median: Strategic Litigants and Legal Argument." Working Paper, UNC-Chapel Hill.

McIntosh, Wayne, Michael Evans, and Cynthia Cates. 2005. "Only Words or Data? Assessing the Relative Policy Positions in Supreme Court Briefs and Opinions." Working Paper, University of Maryland.

TABLE 1: Example Texts and Their Transformed Scores

Text	Words used in text
R1	A A B C D
R2	A B B C E
T1	A B B E E
T2	A B C C
T3	A B D E
T4	A C C D D
T5	A B B C C E

	R1	R2	T1	T2	T3	T4	T5
Raw Score	.367	.633	.733	.500	.500	.267	.611
Our Transformed Scores	0	1	1.38	.500	.500	-.38	.92
LBG Transformed Scores ***(Scoring T1-T5)***			1.39	.43	.43	-.53	.89
LBG Transformed Scores ***(Scoring T2-T5)***				.62	.62	-.52	1.16