Reply to Benoit and Laver

Lanny W. Martin

Department of Political Science, Rice University,
PO Box 1892, MS 24, Houston, TX 77251-1892
e-mail: lmartin@rice.edu

Georg Vanberg

Department of Political Science, University of North Carolina, Chapel Hill,
Chapel Hill, NC 27599-3265
e-mail: gvanberg@unc.edu (corresponding Author)

We appreciate the positive reception of our transformation by Benoit and Laver (hereafter, BL), and we are grateful that they have incorporated it into the Wordscores package. Because their comment highlights a fundamental difference between the Martin-Vanberg (MV) and Laver-Benoit-Garry (LBG) approaches that is critical to the choice among transformations, we offer some brief comments that will allow users to make an informed decision regarding the appropriate use of the transformations. The central issue concerns comparisons between reference and virgin texts. As BL point out, researchers will often be interested in making such comparisons, and the LBG and MV transformations can yield substantially different results. In light of these differences, BL's primary suggestion is to focus analysis on the raw scores, which can be obtained for reference as well as virgin texts. We wholeheartedly agree with this prescription. In fact, it is precisely a concern for faithfully reporting the raw score information, while making it more intuitive, that motivates the MV transformation. As we show below, the MV transformation accurately reflects all and nothing but the information contained in raw scores. Therefore, "users [who] get eye strain" by looking at raw scores can safely substitute MV scores and be confident that the information provided is equivalent. The same will typically not be true of LBG scores.

As BL state, at the core of the LBG transformation is the assumption that the dispersion of virgin and reference texts is the same. Therefore, the transformation adjusts the variance of virgin text scores to equal the variance of reference scores. To compare virgin and reference texts, scholars generate LBG scores for the virgin texts and compare these to the exogenously assigned reference text scores. Note that this implies that in comparing reference and virgin texts, the LBG approach ignores the raw scores of reference texts. Although BL do not investigate the assumption of constant dispersion further, it is likely to be unwarranted in many circumstances. Consider election manifestos. Surely the dynamics of party competition affect not just the *location*, but also the *spread*, of party positions over time, even in adjacent elections. Parties at the ideological extremes in one election may decide to moderate their positions at the next election. Conversely, centrist parties may become more polarized. In either case, the constant variance assumption is inappropriate. And while we agree that "it is hardly uncommon to make strong assumptions about similarities in variances in the practice of applied statistical research," we would add that it

is also not uncommon for researchers to make seriously mistaken inferences when such assumptions are violated in their data. This is why careful analysts only make assumptions that are either empirically justifiable or necessary to make intractable problems tractable. The constant dispersion assumption is obviously not necessary for tractability (since the MV transformation does not require it), and we believe that any researcher who subscribes to the notion of rational, adaptive politicians would (and should) view it as theoretically suspect. Moreover, as we show below, when coupled with the exclusion of reference text raw scores, this assumption distorts the raw score information. This result conflicts with BL's own advice to place primary emphasis on analyzing raw scores.

The logic of Wordscores is to "score" texts mechanically using the dictionary created from the reference texts. It is precisely the simplicity of this approach, which eliminates the need for close analysis of meaning that marks LBG's insight in creating Wordscores. To compare texts, what a researcher therefore wants to know is how the word usage in one text compares to that in another *as judged by the dictionary.* As BL point out, this means that reference and virgin texts should be compared by considering the raw scores assigned to each. However, because raw scores tend to be "bunched," it is helpful to stretch them to a more intuitive metric, while preserving all of the information in the raw scores. This is what the MV transformation does: It takes the raw scores—those of virgin and reference texts—and stretches them linearly. As a result, comparisons based on MV scores lead to *exactly* the same conclusions as comparisons based on raw scores. It is easy to illustrate this using BL's examples.

Consider first the U.K. election example. Suppose an analyst follows BL's advice of looking at untransformed raw scores (see Table 1). What would such an analyst conclude? The raw scores indicate that the Conservative manifesto in 1997—raw score of 10.74—expresses a more moderate position than the Conservative manifesto in 1992—raw score of 11.28. Similarly, Labor's manifesto is more centrist in 1997, moving from a raw score of 9.51 to a score of 10.40. Finally, the raw scores indicate that the Liberal Democrats' manifesto is more centrist in 1997, having moved from a raw score of 9.98 to a score of 10.22. The MV transformation accurately reflects these scores. In contrast, the LBG transformation—which compares the 1997 LBG score to the 1992 reference score while *ignoring* the information contained in the raw score of the 1992 manifesto—concludes that the Liberal Democrats have moved to the left. In other words, an analyst looking at raw scores would draw the same conclusions as an analyst considering MV scores, whereas an analyst using LBG scores would arrive at a different result.[1]

Next, consider the example presented by BL in Table 2. BL note that the MV transformation is sensitive to the selection of reference texts; in particular, that switching from reference texts R1 and R2 to R3 and R4 makes the MV-transformed scores of some virgin texts, notably V6 and V7, appear more "extreme" despite the fact that only "noninformative" content (additional C's) has been added to the reference texts. To take V6 as an example, the MV-transformed score jumps from -1.375 (with R1 and R2 as reference texts) to -3.25 (with R3 and R4 as reference texts). In contrast, the LBG-transformed score of V6 remains right around -0.62. Is this sensitivity, as BL argue, a problem? Once

---

[1]Without further explanation, BL suggest that to judge the LD's 1997 position in relation to its 1992 position, the 1997 raw score should be compared to the mean of all reference scores rather than to the 1992 LD raw score. Such a comparison may allow an analyst to place a text relative to the center of the text distribution, but it is not relevant when trying to assess the relationship between two particular texts. To do this, the appropriate comparison is between the raw scores of the two texts in question. Finally, although it is interesting that expert surveys indicate an LD move to the left, this does not change the fact that the raw scores—which BL purport to prefer—point in the other direction.

again consider an analyst who follows BL's advice of comparing only raw scores. These scores (see Table 2) indicate that V6 is located 1.375 times as far to the left of R1 as R2 is to the right of R1. Similarly, V6 is located 3.25 times as far to the left of R3 as R4 is to the right of R3. That is, V6 looks more extreme compared to R3 and R4 than compared to R1 and R2. If our analyst wants to translate this information onto a scale on which R1 represents a position of 0 and R2 represents a position of 1 (the point of origin, when we use R1 and R2 as reference texts), she would need to place V6 at −1.375 to capture the raw score information. If, on the other hand, she identifies R3 with a position of 0 and R4 with a position of 1, she would need to place V6 at the position −3.25. These are exactly the positions assigned by the MV transformation.

The key to these changes is to recognize that in the Wordscores procedure, adding additional C's to each reference text is not adding noninformative content. Because "C" is shared across reference texts, Wordscores treats it as a "moderate" word. As a result, texts that exclude this word will appear to be more extreme. There is, of course, a complication. The beauty and power of Wordscores is that it simply scores all words, allowing research-ers to analyze large amounts of texts, even in unknown languages. As BL point out, one drawback is that noninformative words that carry no ideological content are included in the analysis. Moreover, such words tend to have significant overlap between texts and are therefore likely to be assigned a centrist score—leading to the bunching of raw scores. We share BL's concern for this problem, which they attempt to address by assuming equal dispersion of reference and virgin texts and using the LBG transformation to "inflate" the variance of virgin text raw scores. However, rather than distorting the raw score informa-tion through a transformation based on a strong and suspect assumption, a more appro-priate solution might be to eliminate "nondiscriminating" words from the analysis. This could even be done within Wordscores through an option that allows users to exclude words above a specified degree of overlap. An analyst who is concerned that noninforma-tive, overlapping words are distorting the results would then be in a position to investigate the robustness of findings to restricting analysis to increasingly "discriminating" words.

LBG's Wordscores program offers an exciting opportunity to engage in systematic content analysis on an unprecedented scale. At stake in the current discussion is the question of how best to take advantage of this potential. Like BL, we believe that, where possible, looking at untransformed raw scores is the first best solution to interpreting the results of a Wordscores analysis. Moreover, we agree that where a "magnifying glass" is needed in order to interpret these scores on a more intuitive metric, any transformation "neither can nor should, in a strict sense, generate any *new* information." Fortunately, unlike the LBG transformation, which adds information (or perhaps misinformation) through its constant variance assumption, the MV transformation satisfies this criterion. Whenever an analyst is interested in comparing across virgin and reference texts, using the MV transformation will faithfully represent *all and nothing but* the information contained in raw scores.