Appendix for:

LeBel, E. P., & Campbell, L. (2009). Implicit partner affect, relationship satisfaction, and the prediction of romantic breakup. *Journal of Experimental Social Psychology*, *45*, 1291-1294. DOI: 10.1016/j.jesp.2009.07.003 [PDF]

[Partner NLT Scoring Algorithm (SPSS Syntax)]

Supplementary Materials

The following discussion summarizes the ongoing debate concerning the optimality of the various scoring algorithms for the Name-Letter Task (NLT) as traditionally applied to measure implicit self-esteem (ISE) (see LeBel & Gawronski, 2009 for more details). This debate has yet to be extended to the romantic partner domain, and thus the goal of this appendix is to summarize the relevant issues (as discussed in the ISE literature) so that interested scholars can apply this knowledge when using the NLT to assess implicit partner affect (IPA).

Kitayama and Karasawa (1997) introduced the first NLT scoring algorithm which involves centering first and last name initial letter ratings around those letter's respective baseline ratings (B-algorithm). Letter baselines are calculated by averaging letter ratings from individuals whose initials do *not* include that letter. As in all algorithms, first and last name preference scores are then averaged to form an index of ISE. In the case of IPA, participants' *partner's* first and last name initial letter ratings are centered around those letter's respective baseline ratings (which in the case of IPA, involves averaging letter ratings from individuals whose own *and* partner's initials do *not* include that letter) (DeHart, Pelham, & Murray, 2004). Although this procedure systematically controls for baseline letter favorability, a major drawback is that the algorithm does not control for individual differences in response tendencies of letter ratings, such as rating all letters relatively high or low, individual differences in positive or negative affect (Watson, 1988), or transient mood states (Schwarz, 1990). Hence one cannot conclude that NLT scores produced by this procedure *uniquely* reflect ISE or IPA.

In order to overcome this limitation, independent researchers in the ISE domain have recently proposed four alternative scoring algorithms. However, attempts at developing a scoring algorithm that controls for both of the aforementioned confounds in a psychometrically and theoretically sound manner has proven to be challenging.

One such algorithm starts with the basic structure of the B-algorithm but then additionally divides the scores by each person's mean ratings of *all* letters (D-algorithm: Gawronski, Bodenhausen, & Becker, 2007). Although this algorithm appears to control for individual response tendencies, upon closer scrutiny it becomes apparent that the procedure introduces a built-in negative relation between response tendency and NLT score, such that the more a person uses the high end of the scale, the lower their resulting score. Hence, the algorithm does not properly control for this confounding factor. In addition, the D-algorithm has been found to have poor psychometric properties, in terms of distribution of scores, outlier production, and relatively low internal consistency (LeBel & Gawronski, 2009).

Another algorithm proposed to control for both confounds, involves first z-transforming each participant's letter ratings within-person (each letter rating centered around each person's mean of *all* ratings divided by his or her own letter ratings standard deviation) (De Raedt, Schact, Franck, & De Houwer, 2006). Then, z-transformed letter preferences for one's initials are centered around their respective z-transformed letter baselines (Z-algorithm). This scoring algorithm is sub-optimal for at least three reasons. First, this approach removes some of the valid ISE variability because the within-person standardization involves subtracting from each letter rating the mean of *all* letter ratings which include the positively biased letter ratings of one's initials. Second, the algorithm distorts scores such that (all else being equal) an individual whose ratings are not that variable will have a larger NLT score compared to an individual whose ratings are highly variable, even though this has nothing to do with ISE. Finally, the Z-algorithm is suboptimal because it has been found to have the lowest internal consistency among scoring algorithms ($\alpha = .33$; LeBel & Gawronski, 2009).

Yet another proposed algorithm uses a unique regression-based approach to systematically control for baseline favorability and individual response tendencies (R-algorithm: Albers, Rotteveel, & Dijksterhuis, 2009). Regression is used to estimate the precise amount by which initial letter ratings are confounded by baseline letter favorability and individual response tendencies (mean of all non-initial letter ratings). Scores are computed by centering raw initial letter ratings around the weighted non-initial letter rating mean and then centering around the weighted respective baseline letter rating (respective weights determined by the relevant regression coefficient from the first step; see Albers et al., for more details). Although this approach ingeniously controls for both confounds, the procedure entails two drawbacks: (a) the resulting scores cannot be interpreted relative to zero to determine whether an overall effect was found, and (b) resulting scores may be relatively "unstable" in small samples given that the aggregation of scores is based on sample-level regression coefficients.

A final algorithm proposed involves ipsatizing letter ratings to control the two confounds (I-algorithm: Baccus, Baldwin, & Packer, 2004). First, the mean rating of all non-initial letters is subtracted from each letter rating (within-person). Then, ipsatized initial ratings are centered around the respective ipsatized baselines (baselines are calculated using the same logic as in the B-algorithm). This algorithm successfully controls for both confounds, and does so in a way that avoids the pitfalls associated with the D-, Z-, and R-algorithms. In addition, it exhibits the strongest psychometric properties in terms of five criteria (LeBel & Gawronski, 2009). That being said, one could argue that because it does not control for within-person *letter rating variability*, correlations to other self-report measures may be slightly inflated. Although this is true in a strict sense, we would argue that built-in within-person *SD* corrections will inevitably cause detrimental score distortions (as in the Z-algorithm), and hence the issue is better dealt with by controlling for within-person variability separately. In the present research, the IPA-RS relation holds when the within-person letter ratings *SD* is added to regression equation.

Given the simultaneous consideration of all these issues, we conclude that the I-algorithm is the most optimal NLT scoring algorithm to index ISE and IPA, and thus recommend the algorithm for use in future research to permit accurate theoretical claims concerning ISE and IPA.