Item Variances and Median Splits: Some Discouraging and Disquieting Findings

Victor Bissonnette, William Ickes, and Ira Bernstein
University of Texas at Arlington
Eric Knowles
University of Arkansas

ABSTRACT In his response to our article, Baumeister argued that in real data, the confound between interitem variance and trait extremity is small (typically in the range of 0 and − 2), and that the danger of artifact associated with the application of median splits to interitem variance is not as serious as our first simulation study would lead one to believe When we examined a large body of actual personality data, employing personality scales of average reliability and relatively large samples, we found that the average magnitude of the confound was − 15 However, we also found that even a confound as small as − 03 could be associated with significant differential range restriction of the trait scores within subsamples produced by the median split (MS) technique We note that several factors, not just the magnitude of the interitem variance/trait extremity confound, must be considered when assessing the danger of artifact associated with the MS technique We again conclude that researchers should use the moderated multiple regression (MMR) technique in preference to the MS technique when testing for moderating effects in personality research

In his response to our article, Baumeister sought to defend the median split (MS) technique as a viable method for testing whether item variance on a trait measure operates as a personality moderating variable (Baumeister, 1990) He noted that, in our first simulation study

Send correspondence to Victor Bissonnette, Department of Psychology, Box 19528, University of Texas at Arlington, Arlington, TX 76019-0528

Journal of Personality 58 3, September 1990 CCC 0022-3506/90/$1 50

(Bissonnette, Ickes, Bernstein, & Knowles, 1990), the correlations between interitem variability and trait extremity were found to be greater than those typically reported by researchers analyzing actual personality data. The smaller confounds characteristic of real data are viewed by Baumeister (1990, Baumeister & Tice, 1988) as typically small or negligible, and not a serious threat to the validity of conclusions based on application of the MS technique.

We agree with Baumeister that it is important to examine actual data to determine the degree of confound between interitem variance and trait extremity, and to assess the danger of artifact posed by this confound. We also agree that when there is *no* correlation between interitem variance and trait extremity ($r = .00$), both the MS technique and the moderated multiple regression (MMR) technique effectively control Type I error rates, as indicated by the results of our second simulation study.

We note, however, that in our first simulation study, we found that the Type I error rate of the MS technique, when applied to interitem variance, varied as a function of a number of factors. For example, as sample size increased, and as the reliability of the personality measure increased, so did the Type I error rate associated with the MS technique. By implication, these findings suggest that the MS technique should result in differential range restriction of the trait scores within the subsamples that it produces (*a*) when the sample size is relatively large, (*b*) when the personality measure used is highly reliable, or (*c*) when both of these conditions occur. Thus, even modest levels of confound (between 0 and −.2) could be expected to result in significant differential range restriction of the subsamples produced by the MS technique when the sample sizes are relatively large and the personality scales are fairly reliable.

To test this prediction, we examined the results of applying the MS technique to a large body of personality data collected in departmental pretestings during previous semesters (fall 1988 through spring 1990). These are the sort of data that are commonly reported in convergent, divergent, and criterion validity studies of personality trait measures. The results of our analyses are presented in Table 1, the first three columns of which specify the personality scale used, the sample size (n), and the scale's reliability coefficient (α).[1]

1 We note that Baumeister and Tice (1988) suggested the use of 9- or 10-point response alternatives when using "metatraits." Because we did not collect data spe-

The fourth column, labeled *r*, lists the interitem variability/trait extremity correlation for each scale Baumeister proposed that, in real personality data, this correlation typically falls between 0 and − 2 Consistent with Baumeister's proposition, of the 24 correlations reported, 18 out of 24 (75%) fell within this range, with a mean correlation of − 15 None of these correlations was positive

It should be noted, however, that the generally low magnitudes of these correlations in no way ensured that the confound between item variance and trait extremity could be dismissed as "not serious " On the contrary, 21 of the 24 correlations (87 5%) between item variance and trait extremity were statistically significant, with 14 of the correlations significant beyond the 001 level

Of greater relevance, however, is the question of whether these modest levels of confound can be associated with significant differential range restriction in the subsamples of trait scores produced by the MS approach In our previous article, we argued that extreme high and extreme low *trait scores* will tend to be associated with less interitem variance and will therefore be overrepresented in the low interitem variance (low IIV) subsamples Thus, we would predict that the *trait score variance* for the low IIV subsamples will be reliably greater than the trait score variance for the high IIV subsamples, indicating the presence of significant differential range restriction

The fifth and sixth columns in Table 1 present the standard deviations of the trait scores within each of the high and low IIV subsamples In 20 of the 24 cases (83 3%), the low IIV subsample exhibited significantly more *trait score variance*, as revealed by the *F* ratios in the last column of Table 1 Moreover, these *F* ratios were significant in 14 of the 18 cases (77 8%) in which the interitem variance/trait extremity confound was between 0 and − 2

These data indicate that when researchers use standard personality scales of average reliability with relatively large samples, even levels of interitem variance/trait extremity confound as small as − 03 can

---

cifically for this investigation, the more common 4- and 5-point response alternative formats were analyzed In a related computer simulation study not reported in our previous article, we found that the Type I error rates associated with the MS technique did not differ a great deal between data representing 5 alternatives and those representing 10 alternatives This previous simulation also supported Baumeister and Tice's hypothesis that the danger of artifact associated with a 2-point response format would be much greater than that associated with a 10-point format

Table 1
Summary of Findings from Pretest Data

Pretest sample	N	α	r	s[a]	s[b]	F
Spring 1990 [c]						
Self-Esteem[d]	543	86	− 10*	6 46	6 01	1 16
Shyness[e]	546	82	− 27****	7 68	5 70	1 81****
Interaction Anxiety[f]	507	89	− 17****	11 71	9 71	1 45**
Fall 1989 [c]						
Private Self-Consciousness[g]	703	70	− 11**	6 15	5 09	1 46***
Public Self-Consciousness[g]	742	77	− 16****	5 78	4 65	1 55****
Social Anxiety[g]	750	75	− 27****	6 09	4 22	2 08****
Shyness[e]	726	83	− 19****	7 96	6 21	1 64****
Interaction Anxiety[f]	736	89	− 09*	12 02	10 51	1 31**
Spring 1989 [h]						
Private Self-Consciousness[g]	465	62	− 10*	4 35	3 75	1 35*
Public Self-Consciousness[g]	494	80	− 14**	3 86	3 42	1 27*
Social Anxiety[g]	493	72	− 22****	4 17	3 34	1 56***
Masculinity[i]	458	88	− 00	8 35	8 77	0 90
Femininity[i]	465	85	− 08	7 77	7 40	1 10
Achievement[j]	476	78	− 03	7 94	7 05	1 27*
Shyness[e]	481	85	− 16***	6 33	4 71	1 81****
Fall 1988 [k]						
Shyness[e]	742	76	− 21****	5 70	4 20	1 84****
Sociability[e]	751	70	− 21****	3 22	2 53	1 62****
Private Self-Consciousness[g]	724	65	− 10**	4 71	3 96	1 42***
Public Self-Consciousness[g]	754	76	− 10**	4 01	3 56	1 27**
Social Anxiety[g]	756	70	− 26****	4 14	3 17	1 70****
Fantasy Ideation[l]	743	70	− 18****	5 96	4 72	1 59****
Empathic Concern[l]	749	78	− 12***	4 45	4 16	1 14
Perspective Taking[l]	746	77	− 15****	5 44	4 32	1 60****
Personal Distress[l]	738	76	− 13***	5 24	4 41	1 42***

*$p <$ 05
**$p <$ 01
***$p <$ 001
****$p <$ 000
a Low interitem variance
b High interitem variance
c Measured using a 5-point Likert response format
d Rosenberg, 1979
e Cheek & Buss, 1981
f Leary, 1983

be associated with significant differential range restriction of the trait scores within the subsamples created by the MS technique If scores on a relevant criterion are then correlated with the trait scores within subsamples evidencing differential range restriction, the trait-criterion correlation within the low IIV subsample will be exaggerated, whereas the trait-criterion correlation within the high IIV subsample will be attenuated This would be true *both* in cases where the subsamples are used to test for a moderating effect of interitem variance, and in cases where one subsample is discarded in preference to the other In the words of Wherry (1984), "Due to prior selection of cases on one of the two variables to be correlated, the resulting correlation coefficient may become *either spuriously high or spuriously low* relative to the value obtained by using the entire sample" (p 49, italics ours, see example on p 50)

The central issue is not so much the size of the correlation between trait extremity and interitem variance, rather, the central issue is whether this confounding, *however small in magnitude*, leads to differential range restriction of the trait scores within the subsamples produced by the median split approach The present data indicate that confounds between 0 and − 2 can lead to significant differential range restriction under conditions implicated by the first of our simulation studies

CONCLUSION

Is the danger of artifact associated with the application of median splits to interitem variance a "common malady" or a "rare and exotic disease?" It depends The results of our first simulation study demonstrated that the Type I error rate of the MS approach, when applied to interitem variance, is a function of a number of factors, such as scale reliability and sample size For example, as sample size increases, and as reliability increases, the potential for artifact also increases

---

g Feningstein, Scheier, & Buss, 1975

h Measured using a 4-point Likert response format

i Bem, 1974

j Spence & Helmreich, 1983

k Shyness, Sociability, Private Self-Consciousness, Public Self-Consciousness, and Social Anxiety were measured using a 4-point Likert response format Fantasy Ideation, Empathic Concern, Perspective Taking, and Personal Distress were measured using a 5-point Likert response format

l Davis, 1983

Complementing these findings from the simulation studies, the present examination of real personality data indicated that when standard personality scales were administered to relatively large samples, significant differential range restriction was associated with the MS approach 83.3% of the time. Moreover, such differential range restriction was found in 77.8% of the cases in which the interitem variance/trait extremity confound was between 0 and −.2.

Baumeister correctly interpreted our first simulation study as indicating that researchers should beware the danger of artifact when the interitem variance/trait extremity correlation is large. However, he neglected to note that the rate of Type I error associated with the application of the MS technique to interitem variance is *multiply determined by a number of factors* that include sample size, scale reliability, and the size of the trait-criterion correlation. Instead of acknowledging this complexity, Baumeister proposed a simple but insufficient rule-of-thumb, i.e., that the magnitude of the interitem variance/trait extremity correlation can serve as a single, reliable indicator of the danger of artifact associated with applying median splits to interitem variance.

The results we have reported here and in our previous article lead us to a different conclusion. We conclude that researchers should not rely on the magnitude of the moderator/trait extremity confound when assessing the danger of artifact associated with the use of median splits to test for moderator effects in personality research. We suggest that the danger of artifact is instead a more complex function of a number of parameters that are likely to vary from one personality study to the next. Specifically, even when the interitem variance/trait extremity confound is low, the MS approach is susceptible to range-restriction artifact when scale reliability is high, when sample size is large, or when both of these conditions apply. We further note that some of the conditions that reduce the danger of artifact associated with median splits of interitem variance (e.g., small sample size, low scale reliability, and weak trait-criterion correlation) are conditions that many researchers deliberately seek to *avoid* when designing a personality study.

For these reasons, we again recommend against using the MS technique to test for personality moderating effects. Instead, we recommend using the MMR approach, which our simulation studies have shown (*a*) to control Type I error rate across all tested levels of sample size, scale reliability, and trait-criterion correlation, and (*b*) to be more statistically powerful than the MS technique. The MMR analysis, whether used alone or as a check on the MS analysis, should enable researchers to maximize the probability of detecting genuine moderating effects, and minimize the probability of reporting spurious ones

REFERENCES

Baumeister, R (1990) Item variances and median splits Some encouraging and reassuring findings *Journal of Personality*, **58**, 589–594

Baumeister, R , & Tice, D (1988) Metatraits *Journal of Personality*, **56**, 571–598

Bissonnette, V , Ickes, W , Bernstein, I , & Knowles, E (1990) Personality moderating variables A warning about statistical artifact and a comparison of analytic techniques *Journal of Personality*, **58**, 567–587

Bem, S (1974) The measurement of psychological androgyny *Journal of Consulting and Clinical Psychology*, **42**, 155–162

Cheek, J , & Buss, A (1981) Shyness and sociability *Journal of Personality and Social Psychology*, **41**, 330–339

Davis, M (1983) Measuring individual differences in empathy Evidence for a multidimensional approach *Journal of Personality and Social Psychology*, **44**, 113–126

Feningstein, A , Scheier, M , & Buss, A (1975) Public and private self-consciousness Assessment and theory *Journal of Consulting and Clinical Psychology*, **43**, 522–527

Leary, M (1983) Social anxiousness The construct and its measurement *Journal of Personality Assessment*, **47**, 66–75

Rosenberg, M (1979) *Conceiving the self* New York Basic Books

Spence, J , & Helmreich, R (1983) Achievement-related motives and behaviors In J T Spence (Ed), *Achievement and achievement motives Psychological and sociological approaches* (pp 7–74) San Francisco Freeman

Wherry, R (1984) *Contributions to correlational analysis* Orlando Academic Press

Manuscript received February 28, 1990, revised March 12, 1990