LITOTES RESPONSE SET

Willem K.B. Hofstee, Feb. 2010
With many thanks to A.A. Jolijn Hendriks

Recently, at a ceremonial occasion, I carried out a case study of one of my colleagues' personality. Five other colleagues scored him on the 100 behavioral items, phrased in the third person singular, of the Five Factor Personality Inventory (FFPI; see, Hendriks, Hofstee, & De Raad, 1999), using its 5-point Likert scale running from "Not at all applicable" to "Entirely applicable", for short, [NO, no, ?, yes, YES], or [1….5]. The individual assessors were instructed to predict the median (see, Hofstee, 2009) per item of their joint ratings, to discourage idiosyncratic assessment; a symbolic prize was awarded to the winner of this recursive bet. This note is not to report on the substantive findings of the study, but to document a remarkable tendency in the response patterns, that has hitherto escaped students of personality assessment; apparently, it needed the kind of close inspection of raw data that is seldom carried out by researchers.

Agreement among the assessors was perfect on 8 items[1)]; in all these cases but one, a string of five NO's occurred, the exception being a string of five midpoint responses. All 7 unanimous NO's were on socially undesirable items. So, it looks as if our target colleague's personality was primarily assessed through litotes[2)], in this case, affirmation of socially desirable traits by emphatic negation of their contraries.

On further inspection, it appeared that all 5 assessors gave more NO's than YES's, even though the ratios differ ($\chi^2 = 12.7$, df = 4, $p < .05$):

	assessors					
	A	B	C	D	E	Σ
YES	19	4	12	11	16	62
NO	31	31	21	20	24	127

An interpretation in terms of nay-saying response set would be incorrect: Together, the assessors assigned slightly more yes's than no's. The total distribution of the 500 responses was:

YES	yes	?	no	NO
62	109	103	101	127

Finally, the 100 item medians were distributed as follows:

YES	yes	?	no	NO
10	22	20	23	25

The effect might be due to the target person, the assessors, the use of informant ratings rather than self-ratings, or the instrument. To check some of these hypotheses, I inspected the data of the Dutch normative sample on the FFPI, consisting of self-ratings by 2494 individuals (Hendriks et al., *ibid.*). The mean counts for the response categories were:

YES	yes	?	no	NO
9	28	26	23	14

There are fewer NO's in these self-ratings, but their preponderance over YES's remains pronounced, which means that the first three explanations are not sufficient. The question remains whether one would expect the litotes effect in other questionnaires, or whether it comes about specifically in the FFPI. For, from the heydays of the response set literature (see, e.g., Wiggins, 1973) one is left with the impression that acquiescence rather than NO-saying is the phenomenon to be expected.

First, one should distinguish between differential and global acquiescence. Differential acquiescence, the tendency for one individual to find more items applicable than another, is an established phenomenon; for example, the FFPI Acquiescence scale[3)] has a high reliability. However, differential acquiescence is the same as differential nay-saying. Global acquiescence, a general tendency to find items more applicable than not, is quite a different thing. On the FFPI, the mean score on the Acquiescence scale in the normative sample is slightly negative. So, the question is whether the FFPI is atypical in this respect.

The FFPI has been constructed as a balanced set of items, in terms of positive and negative item loadings on the five factors. The distribution of the item social desirabilities, on a [1…5] scale, is shown in Figure 1. There appear to be 49 desirable and 51 undesirable items; the mean social desirability of the 100 items is 3.0. If anything, the number of clearly undesirable (< 2.0) items is lower than the

number of clearly desirable (≥ 4.0) items, namely, 7 *versus* 24. Thus the litotes effect is not elicited by an overrepresentation of clearly undesirable items.

Figure 1

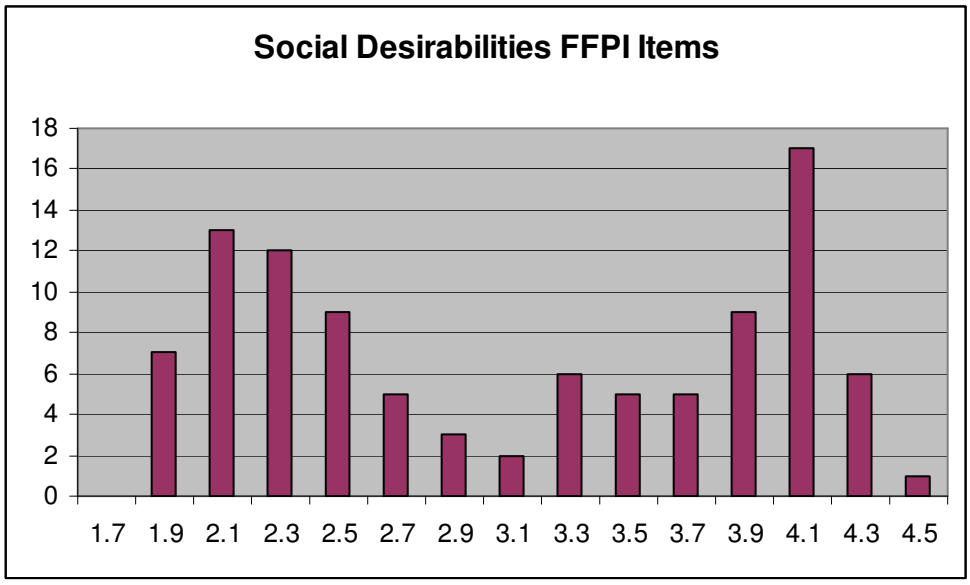
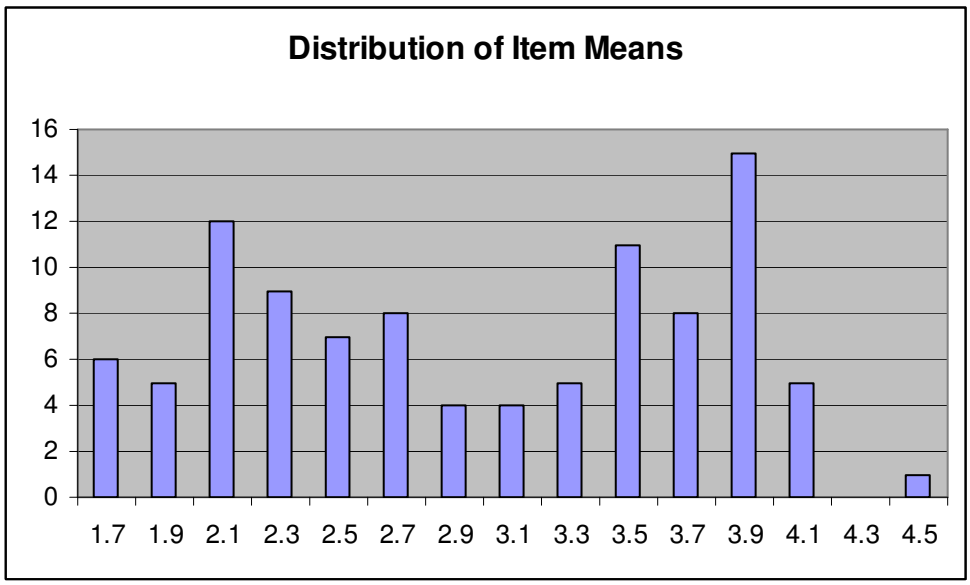


Figure 2 shows the distribution of the item means in the normative sample. Their grand mean is also at the scale midpoint of 3.0, but the number of little endorsed (< 2.0) items surpasses the number of greatly endorsed (≥ 4.0) items by 11 *versus* 6, running counter to the ratio of the tail social desirabilities, and illustrating the litotes effect.

Figure 2



To complete the picture, I split the FFPI into socially desirable and undesirable items, and used the normative sample data to calculate the percentage of YES's and NO's:

	YES	NO
Desirable items	16.2	3.2
Undesirable items	2.4	26.1

The pattern in this fourfold table may be easier to read when contrasting emphatic desirable (YES to a desirable item or NO to an undesirable item) with emphatic undesirable responses:

	Emphatic Desirable Responses	Emphatic Undesirable Responses
Desirable items	YES: 16	NO: 3
Undesirable items	NO: 26	YES: 2

There are two main effects: more emphatic responses to undesirable items, and of course, more desirable than undesirable responses. But there is also an antagonistic interaction: both the desirable and the undesirable items elicit relatively many NO responses. The interaction is even more pronounced in our $N = 1$ study:

	Emphatic Desirable Responses	Emphatic Undesirable Responses
Desirable items	YES: 12	NO: 4
Undesirable items	NO: 22	YES: 1

The response pattern may be described as an overrepresentation of emphatic denials at the expense of emphatic affirmations, especially but not exclusively on socially undesirable items. In classical terms, it is a combination of nay-saying and extreme response set, but "litotes tendency" is more descriptive. Clearly, in an unbalanced questionnaire the pattern would come out differently; with very one-sided questionnaires *qua* social desirability, it could not even be observed. But that is not to say that the phenomenon would not exist, and least of all, that the FFPI is atypical in this respect.

One further characteristic of the FFPI deserves attention: In none of the 100 items, negative formulations were admitted (e.g., “not”, “no”, “un-“), the reason being that negatively worded items are confusing, especially when they would be found “not applicable”, leading to double negations. To the extent that negative items occur in other questionnaires, one would expect that fewer NO’s would occur, because of confusion. Again, however, the FFPI is exemplary rather than atypical in this respect.

The reader may speculate about mechanisms that bring about litotes set, and individual differences in that respect. For example, would the phenomenon hold in other languages and cultures? Is it specific to personality assessment, or does it show up in other uses of the Likert scale as well? What is the cognitive mechanism, running counter to the more familiar confirmation bias (e.g., Davies, 2003)? Do the individual differences in litotes tendency have any external validity – as was hypothesized about acquiescent response set, but not consistently verified?

For practical as opposed to speculative purposes, litotes set is a nuisance variable, as it biases the response scale to some extent. If respondents say “yes” to an item and “NO” to its opposite, they contribute to an asymmetry in the scale. This kind of asymmetry is especially disturbing in the construction phase of a questionnaire: For example, differential acquiescence has been known to result in correlations between opposite items that were closer to 0 than –1. (In the application phase, acquiescence is taken care of by using a balanced set of items).

Correcting for litotes set might be achieved through symmetricizing the scores distribution by weighting the responses. Take the distribution of the self-ratings in the normative sample and correct by averaging pairwise:

	YES	yes	?	no	NO
Percentage	9	28	(26)	23	14
Corrected	11.5	25.5	(26)	25.5	11.5
Weight	1.28	.91		1.11	.82

Note that the weighting also automatically corrects for both global and differential acquiescence. However, it is rather unwieldy. One may therefore consider using the correction for nay-saying, which consist of a simple additive constant, as an approximation: It works to some extent because there are far fewer NO’s to socially desirable items than to undesirable items. One should realize, however, that acquiescence is not the real thing, and that it does not correct for a social desirability bias that is the net effect of more NO’s to undesirable than desirable items. Finally, a more radical solution would be to wipe out both global and differential litotes set by taking YES’s and yes’s together, and NO’s and no’s. In the normative data, the [Yes, ?, No] distribution is [37, 26, 37], which happens to be perfectly symmetrical.

References

Davies, M.F. (2003). Confirmatory bias in the evaluation of personality descriptions: Positive test strategies and output interference. *Journal of Personality and Social Psychology, 85*(4), 736-744.

Hendriks, A.A.J., Hofstee, W.K.B., & De Raad, B. (1999). The five-factor personality inventory (FFPI). *Personality and individual Differences, 27,* 307-325.

Hofstee, W.K.B. (2009). Promoting intersubjectivity: a recursive-betting model of evaluative judgements. *Netherlands Journal of Psychology, 65,* 2-9.

Wiggins, J.S. (1973). *Personality and Prediction.* Reading, Mass.: Addison-Wesley.

Footnotes

[1)] This may seem little, but the probability of perfect agreement on random responding is only $.2^4 = .0016$. More generally, agreement was high: The average distance per item between an individual rating and the median rating ranged from .32 to .54 for the 5 assessors, whereas 1.56 would be expected on random responding.

[2)] Webster defines litotes as understatement for effect (Ex.: not a few regrets). The effect is closer to an overstatement, which is why the term fits the emphatic denials in the present context.

[3)] The FFPI Acquiescence score is the mean per individual over 60 (30 pairs of near opposites) items, so that a score at the scale midpoint means null acquiescence.