Behavioral variability measured using Likert and frequency based response formats

Rizwan Khan

Michael D. Biderman

University of Tennessee at Chattanooga

Authors' Note: Correspondence regarding this article should be sent to Michael Biderman, Department of Psychology / 2803, U.T. Chattanooga, 615 McCallie Ave., Chattanooga, TN 37403. Tel.: (423) 425-4268. Email: Michael-Biderman@utc.edu

Paper accepted for presentation at the 28th Annual Conference of The Society for Industrial and Organizational Psychology, Houston, TX. 2013.

Poster

TITLE
Behavioral variability measured using Likert and frequency based response formats

ABSTRACT
Likert and frequency based response formats were used to measure inconsistency in self-report of personality and variability in levels of personality. Inconsistency contributed incremental validity over conscientiousness in prediction of GPA. Variability in levels of personality from the frequency based response format correlated positively with a measure of ADHD.

PRESS PARAGRAPH
Most personality research and practice has focused on levels of personality traits. This research focused on variability associated with personality. Inconsistency of self-report from a frequency based response format and a traditional Likert format each contributed incremental validity over conscientiousness in prediction of GPA. Variability of personality measured from the frequency based format was positively related to a measure of ADHD. Most analyses suggested that inconsistency of self report and variability of personality are different constructs.

Although most research and practice involving personality questionnaires focuses on the levels of respondents on personality dimensions, there is growing interest in other characteristics measurable from such questionnaires. Probably the most frequently studied of those other characteristics is the variability of respondents' scores on personality scales. Such variability has been considered for many years (e.g., Allport, 1937; Baird, 2006; Fiske & Rice, 1955 Fleeson, 2001). Most of the measures of variability have been based on scale scores, examining variation in such scores across situations, across time within situations, and across both times and situations (e.g., Eid & Diener, 1999; Fleeson, 2001). The variation in scale scores is typically viewed as variation in the actual level of a specific characteristic. Here it will be called state/trait variability.

Although there is much interest in state/trait variability, there is also interest in variability in the report of personality characteristics. Recently, Reddock, Biderman, & Nguyen (2011) and Biderman and Reddock (2012) have shown that inconsistency in self-report defined as the standard deviation of responses to individual Likert items within a dimension is related to both scale reliability and validity. Respondents who were more inconsistent in responding to the items of a questionnaire yielded less reliable scale scores that had lower criterion-related validity than respondents who whose responses were less inconsistent.

Although self-report inconsistency can be assessed within a single experimental setting as the standard deviation of responses to items within a questionnaire dimension, the measurement of state/trait variability typically has required multiple observations of respondents across situations or time periods. Recently, however, a response format has been investigated that has the promise to inform researchers and practitioners on both the level of a personality characteristic possessed by respondents and on state/trait variability of that characteristic from

data gathered within a single setting. As presented by Edwards and Woehr (2007) and Fleischer, Woehr, Edwards, and Cullen (2011) this innovation requires respondents to indicate percentage of time the behavior described by a personality questionnaire item was very inaccurate, was neither accurate nor inaccurate, and was very accurate. By assigning values to each response category, the authors were able to compute derived item scores comparable to Likert responses to each item, thus representing the level of the characteristic, and also a value that potentially represented the variability in the extent to which the behavior represented by the item was an accurate description of the respondent, a measure of state/trait variability. The authors provided evidence that the derived item scores yielded scale scores with high convergent validity with traditional Likert scale scores. They also presented evidence of the validity of state/trait variability estimates for individual personality dimension scores. There is not yet, however, any evidence on the relationship of the measures of variability computed from the frequency based response format with the inconsistency of self-report studied by Reddock et al. (2011) and Biderman and Reddock (2012). The purpose of the present research was to attempt to fill this gap by studying the relationship of state/trait variability as measured using the frequency based response format with self-report inconsistency measured using the within-dimension standard deviations of traditional Likert item responses.

Interestingly, because the frequency based response format yields a single derived “level” value for each item analogous to a single Likert response, the variability of those derived Likert item scores might represent a measure of self-report response inconsistency comparable to the measures computed from actual Likert responses to items. For that reason a second purpose of the present research was to investigate the relationship of self-report inconsistency computed from the derived Likert responses of the frequency based response format with self-report

inconsistency computed from traditional Likert responses.

Based on the work of Edwards and Woehr (2007) and Fleisher et al. (2011) we expected that the scale scores representing corresponding dimensions from the two response formats would exhibit convergent validity. Since variability of behavior is clearly not equivalent to inconsistency in the self-report of that behavior, we did not expect a high correlation between state/trait variability from the frequency based response format and self-report inconsistency from either the traditional Likert responses or the derived Likert item values. We did expect convergent validity of self-report inconsistency measures from the traditional Likert responses and the derived Likert item values.

Since both state/trait variability and self report inconsistency are conceptually distinct from the levels of personality traits, it was of interest to investigate the discriminant validity of the variability measures with the trait level measures. Reddock et al. (2011) found that self-report inconsistency measures were essentially uncorrelated with trait level scores of the Big Five. On the other hand, Fleisher et al. (2011) in their Table 3 reported significant positive correlations between frequency based measures of state/trait variability and levels of agreeableness, conscientiousness, and emotional stability, the only Big Five variables assessed in their study. This study extends these results to include an investigation of the discriminant validity of traditional Likert inconsistency, inconsistency from derived Likert scores, and frequency based state/trait variability with all Big Five dimensions.

Reddock et al. (2011) found that self-report inconsistency contributed incremental validity to prediction of GPA over both cognitive ability and conscientiousness – persons with high inconsistency had lower GPAs. Furthermore, they found that inconsistency moderated the relationship of GPA to conscientiousness, with lower correlation of GPA with conscientiousness

among those who were most inconsistent. Fleisher et al. (2011) found that state/trait variability based on the frequency based response format moderated the relationship of peer reports of performance to agreeableness and conscientiousness. For this study, we investigated the validity of all three measures of variability for prediction of GPA. Based on the results of Reddock et al. (2011) and Fleisher et al. (2011), we expected that both state/trait variability and self-report inconsistency would moderate the relationship of conscientiousness to GPA.

The final purpose of this study was to determine whether a relationship existed between any of the measures of variability and Attention Deficit Hyperactivity Disorder. ADHD is characterized by varying levels of attentive issues and/or hyperactivity among children which may lead into adulthood. Current ADHD research has involved analyzing reaction times during tasks to determine the degree to which inconsistency occurs in the reaction times. It has indicated that individuals who possess ADHD are more likely to be inconsistent in their reaction times (Adams, Roberts, Milich, and Fillmore, 2011; Castellanos & Tannock, 2002).We included a measure of ADHD in the present study in order to investigate the relationship of ADHD to the three measures of variability in personality and its assessment. Based on the previous research involving reaction times and the conceptualization of ADHD as a syndrome notable for variability in behavior we expected positive correlations of the variability measures with the ADHD measures.

Method

Participants

Participants were 175 undergraduates at a southeastern university participating for course credit. Forty-three were male. Mean age was 18.83 (SD=2.04). Percentage of Whites, Black/African-Americans, Hispanics, and Other were 78.4, 12.0, 3.0, and 6.8 respectively.

Measures

Big Five Scales. Each participant completed two IPIP Big Five (Goldberg, 1999; www.ipip.org) questionnaires. Items from the 50-item sample questionnaire were called the Original questionnaire and items from the 100-item sample questionnaire that were not in the 50-item sample were the called Other questionnaire.

Response formats. Respondents filled out one questionnaire using a Likert response format and the other using a frequency based format. The Likert format used a 7 point scale ranging from “Completely Inaccurate” to “Completely Accurate”. The frequency based response format three response categories: “Very Inaccurate”, “Neither Inaccurate nor Accurate”, and “Very Accurate”. Participants assigned a percentage value to each category representing the frequency with which the behavior indicated by the item described the participant, with the restriction that the percentages sum to 100%. The percentages in the three cateogires were multiplied by .01, .04, or .07 to make a derived Likert score for each item ranging from 1 to 7.

Variability. Self-report inconsistency was operationalized for the Likert scale response by computing the standard deviation of responses of each participant to items within a dimension, and averaging those standard deviations across dimensions (Biderman and Reddock, 2012; Reddock, et al., 2011). Self-report inconsistency was operationalized for the frequency based response format in the same way but was based on the derived scores described above.

For state/trait variability, a standard deviation for each item was computed based on the percentage responses for each item and the values .01, .04, and .07 for the response categories. The mean of the individual item standard deviations across the ten items within a dimension was computed to yield a measure of state/trait variability for that dimension, and the mean of the standard deviations from the five dimensions was computed to obtain the overall measure of

state/trait variability analyzed here.

GPA. Cumulative GPA for each respondent was obtained at the end of the semester in which participation in the research occurred.

ADHD. Individuals completed the 25-item Wender Utah Rating Scale (WURS) developed by Ward, Wender, and Reihmerr (1993). The questionnaire consists of a 5 point scale ranging from “Not at all or very slightly” to “Very Much”. A total scale score for each respondent was computed and correlated with the variability measures.

Procedure

The Big Five questionnaires were given to each participant. The order of presentation of the two formats and two versions of the Big Five questionnaire were counterbalanced. One-fourth of the respondents responded to the two questionnaires in each of the combinations of response format and questionnaire. The WURS was administered last.

All participants who responded using the frequency based format to the Original questionnaire and the Likert format to the Other questionnaire were treated as one group of 86 participants. All participants who responded using the frequency based format to the Other questionnaire and the Likert format to the Original questionnaire were treated as a second group of 89 participants. Parallel analyses were conducted on the two groups.

Results

Table 1 presents reliabilities of the Likert Scales, the frequency based scales, and the inconsistency measures. All reliabilities with the exception of the reliability of Likert inconsistency from the Original questionnaire were at least marginally acceptable.

Table 2 presents convergent validity correlations between the frequency based derived scale scores with the Likert scale scores. Correlations between scores measuring the same

dimension were large and positive.

Table 3 presents the correlations among state/trait variability, and the two inconsistency measures. As expected, self-report inconsistency based on traditional Likert scores was positively correlated with inconsistency computed from derived Likert values, but not correlated with state/trait variability. Interestingly, derived inconsistency scores were negatively correlated with the state/trait variability values.

Table 4 presents discriminant validity correlations between the scale scores and variability values. For both inconsistency measures discriminant validity correlations were generally nonsignificant. On the other hand, eight scale scores significantly negatively correlated with the stat/trait variability measure, indicating a general a lack of discriminant validity.

Table 5 presents results of regressions of GPA onto conscientiousness and measures of variability. Two-predictor regressions involving conscientiousness and variability are on the left. On the right are regressions with a product variable added to assess moderation. In order to maximize sample size, the data from both groups were analyzed together with a variable representing the distinction between questionnaires. For both response formats both conscientiousness and inconsistency contributed significantly to the prediction of GPA although the result was only marginally significant for inconsistency based on the traditional Likert response format. The sign of the inconsistency variable was negative, indicating that those with higher inconsistency had lower GPAs. In neither instance was inconsistency a moderator of the conscientiousness-GPA relationship. For both product variables, the sign of the product was negative, indicating that conscientiousness was (insignificantly) less valid for respondents who were more inconsistent.

The state/trait variability measure did not meet the .05 level of significance, and its

coefficient was positive. The product variable, while negative, was not significant. Thus there is no evidence here of a relationship of state/trait variability to GPA.

Table 6 presents the correlations of the measures with the ADHD measure. The inconsistency measures were not related to ADHD scores. However, the state/trait variability measure was positively correlated with ADHD in both groups.

Discussion

The purpose of this study was to compare two conceptualizations of personality variability – variability in actual levels of personality across time and inconsistency in the self-report of personality. We assessed the convergent and discriminant validity of three measures of variability along with their criterion-related validity and their relationships with a measure of ADHD.

First, this study found support for the convergent validity of Big Five scale scores created from the two formats. These uncorrected correlations are substantial in view of the fact that they were across response formats and across questionnaires. This result replicates those of Fleisher et al. (2011) and suggests that the frequency based response format measures personality levels to a degree similar to that from Likert scales.

The results in Table 3 suggest that as expected self-report inconsistency measured from the two formats exhibited some convergent validity, although the relationship was significant for only one format-questionnaire group. And, as might be expected, self-report inconsistency measured from traditional Likert responses and state/trait variability appear to be orthogonal constructs. Surprisingly, though, self-report inconsistency measured from the derived Likert values was negatively correlated with state/trait variability. The result may be an artifact of the response format. Those with the largest state/trait variability values would be those whose

percentage responses were confined to the two extreme categories, ignoring the center. The frequent use of only the outer response categories may have resulted in derived item values that were less variable from item to item than for respondents who responded with percentages in all three categories, resulting in the observed negative correlations. This represents a possible limitation of this response format.

Because variability of behavior is conceptually distinct from level of behavior, it was of interest to examine the discriminant validity of the variability measures with levels of the Big Five dimensions. As shown in Table 4, the self-report inconsistency measures were generally unrelated to levels of the Big Five dimensions suggesting that self-report inconsistency from both response formats is not only conceptually distinct but statistically distinct from level of Big Five dimensions. On the other hand, state/trait variability estimates were mostly negatively related to level. This result is quite different from the results reported in Fleisher et al. (2011) Table 3, where the correlations between measures were positive. We note that our measure of variability was averaged across all Big Five dimensions, while those reported by Fleisher et al. (2011) were dimension specific. At the present time we have no explanation for the correlations involving state/trait variability or the discrepancy between the results found here and those from previous research.

Assessment of criterion-related validity of both measures of self-report inconsistency replicated previous results. Self-report inconsistency was found to be incrementally valid over conscientiousness in prediction of GPA. The results certainly suggest that selection specialists can take advantage of the essentially free measure of self-report inconsistency to increase criterion-related validity. The results for assessment of moderation were less conclusive. Although the signs of the product variables were negative as would be expected if

conscientiousness were less valid for more inconsistent respondents, neither product variable was significant.

In contrast to the findings of Fleisher et al. (2011) there was little evidence for the efficacy of the measure of state/trait variability. It is possible that differences in the criteria between the two studies are an explanation for the difference in result. Their criterion was peer ratings of team member performance, a criterion involving social interactions between the raters and ratees. It is possible that persons with stable personality levels are rated more highly than those who exhibit variability in personality levels. Since GPA is far less dependent on social interactions, it may be that stability of levels was less salient for academic performance.

Since ADHD is characterized by inattentiveness and impulsivity, it was hypothesized that individuals with ADHD would exhibit greater variability. The results suggested that self-report inconsistency is not the kind of variability that is related to ADHD scores. On the other hand, there were significantly positive relationships between state/trait variability and scores on the ADHD scale. This provides support for the notion that the frequency based response format may provide information about behavior that will be useful in understanding characteristics of behavior other than levels on personality dimensions.

Limitations

The small reliabilities of the self-report inconsistency measures may have suppressed the correlations of these measures with other variables, for example, in assessing the discriminant validities of Table 4. On the other hand, the low reliabilities did not prevent both from contributing to validity of prediction of GPA in Table 5.

Future Directions

A clear direction for future research is an assessment of the relationship of the state/trait

variability measure used here with other measures of state/trait variability based on experience sampling techniques (Fleeson, 2001) or other multiple measures of personality. A demonstration of convergent validity of the frequency based measure with other measures would certainly represent an argument for use of the frequency based response format.

References

Adams, Z. W., Roberts, W. M., Milich, R., & Fillmore, M. T. (2011). Does response variability predict distractibility among adults with attention-deficit/hyperactivity disorder? *Psychological assessment, 23,* 427-436.

Allport, G.W. (1937). *Personality: A Psychological Interpretation.* New York: Holt.

Baird, B. M., Le, K., & Lucas, R. E. (2006). On the nature of intra-individual personality variability: Reliability, validity, and associations with well-being. *Journal of Personality and Social Psychology*, 90(3), 512-527.

Biderman, M.D., & Reddock, C. M. (2012). The relationship of scale reliability and validity to respondent inconsistency. *Personality and Individual Differences, 52*, 647-651.

Castellanos, F. X., & Tannock, R. (2002). Neuroscience of attention deficit/ hyperactivity disorder: The search for endophenotypes. *Nature Reviews Neuroscience, 3,* 617–628.

Edwards, B. D., Woehr, D. J. (2007). An examination and evaluation of frequency-based personality measurement. *Personality and Individual Differences, 43*, 803-814.

Eid, M., & Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and social Psychology, 76*, 662-676.

Fiske, D. W., & Rice, L. Intra-individual response variability. *Psychological Bulletin, 52*, 217-250.

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: traits as

density distributions of states. *Journal of personality and social Psychology, 80*, 1011-1027.

Fleisher, M. S., Woehr, D. J, Edwards, B. D, & Cullen, K. L. (2011). Assessing within-person personality variability via frequency estimation: more evidence for a new measurement approach. *Journal of Research in Personality, 45*, 535-548.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe*, Vol. 7 (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.

Ward, M. F., Wender, P. H., & Reimherr, F. W. (1993). The Wender Utah Rating Scale: An aid in the retrospective diagnosis of childhood attention deficit hyperactivity disorder. *The American Journal of Psychiatry*, 150(6), 885-890.

Table 1

Reliabilities of scale scores and inconsistency measures

	Likert		Frequency Based	
	Group 1	Group 2	Group 1	Group 2
Extraversion	.834	.885	.752	.745
Agreeableness	.664	.716	.666	.638
Conscientiousness	.815	.728	.768	.739
Stability	.717	.907	.780	.809
Openness	.638	.821	.745	.729
Self-report Inconsistency	.693	.533	.641	.733
State/Trait Variability			.869	.915

Note. In Group 1 the frequency based response format was used for the Original questionnaire and Likert for the Other questionnaire. In Group 2 the frequency based response format was used for the Other questionnaire and the Likert for the Original questionnaire. Inconsistency was computed as the mean of within-dimension standard deviations of item responses. Variability was computed as the mean of individual-item standard deviations based on percentages assigned to response categories.

Table 2

Correlations between Frequency-based scale scores and Likert scale scores

	Likert Scale Scores									
	E		A		C		S		O	
Frequency based Scale Scores										
E	**.662**[c]	**.741**[c]	.307[b]	.024	.013	.267[a]	-.052	.356[b]	.248[a]	.143
A	.297[b]	.219[a]	**.660**[c]	**.554**[c]	.375[c]	.252[a]	.050	.221[a]	.256[a]	.146
C	.109	.080	.234[a]	.147	**.719**[c]	**.765**[c]	.149	-.003	.261[a]	-.082
S	.026	.172	.094	-.056	.027	.058	**.643**[c]	**.717**[c]	.083	.201
O	.364[b]	.129	.170	.178	.146	.142	.046	.350[b]	**.663**[c]	**.745**[c]

Note. Convergent validities are in boldface. Leftmost entries are from Group 1for which the frequency based response format was used for the Original questionnaire and Likert for the Other questionnaire. Rightmost entries are from Group 2 for which the frequency based response format was used for the Other questionnaire and the Likert for the Original questionnaire.

E= Extraversion; A= Agreeableness; C= Conscientiousness; S= Stability; O= Openness.

[a] $p<.05$ [b] $p<.01$ [c] $p<.001$;

Table 3

Correlations between the Frequency-based and Likert Inconsistency Scores

	Likert Inconsistency	Frequency based Inconsistency	Frequency based Variability
Likert Inconsistency	1	.436[c] | .162	-.145 | -.074
Frequency based Inconsistency		1	-.422[c] | -.437[c]
Frequency based Variability			1

Note: Leftmost entries are from Group 1 for which the frequency based response format was used for the Original questionnaire and Likert for the Other questionnaire. Rightmost entries are from Group 2 for which the frequency based response format was used for the Other questionnaire and the Likert for the Original questionnaire.

[a] $p<.05$ [b] $p<.01$ [c] $p<.001$;

Table 4

Discriminant validity correlations of variability measures with Big Five scale scores

	Likert	Frequency based	State/trait
	Inconsistency	Inconsistency	Variability
Extraversion	.007 | .205	.018 | -.045	-.075 | -.282[b]
Agreeableness	-.032 | .170	.044 | .061	-.443[c] | -.448[c]
Conscientiousness	.224[a] | -.037	.090 | .073	-.505[c] |-.374[c]
Stability	-.102 | .063	-.210 | .105	.105 | -.305[b]
Openness	-.032 | .103	.052 | -.037	-.313[b] | -.386[c]

Note: The left column presents discriminant validity correlations between the Likert scale scores and Likert inconsistency score. The middle column presents discriminant validity correlations between the FB dimensional scale scores and FB inconsistency score. The right column presents discriminant validity correlations between the state/trait variability measures and frequency based scale scores. Leftmost entries are from Group 1 for which the frequency based response format was used for the Original questionnaire and Likert for the Other questionnaire. Rightmost entries are from Group 2 for which the frequency based response format was used for the Other questionnaire and the Likert for the Original questionnaire.

[a] $p<.05$ [b] $p<.01$ [c] $p<.001$;

Table 5.

Moderated regressions of GPA onto conscientiousness and variability measures.

	Main Effects Model		Moderation Model	
	Std'd Coeff	p	Std'd Coeff	p
Likert format self-report inconsistency				
Questionnaire	.109	.159	.101	.193
Conscientiousness	.234	.003	.850	.028
Inconsistency	-.130	.083	.577	.189
C x I			-1.005	.103
R^2	.253		.281	
Frequency based self-report inconsistency				
Questionnaire	.069	.347	.069	.350
Conscientiousness	.214	.004	.390	.276
Inconsistency	-.222	.003	.007	.989
C x I			-.301	.615
R^2	.300		.302	
State/trait variability				
Questionnaire	.052	.489	.052	.488
Conscientiousness	.259	.002	.203	.315
Variability	.146	.078	.140	.099
C x V			.059	.760
R^2	.242		.243	

Table 6

Correlations with ADHD Measure

Variable	Group 1	Group 2
Likert Inconsistency	.006	-.012
Frequency based inconsistency	.033	-.075
Frequency based state/trait variability	.248[a]	.285[b]

Note: In Group 1 the frequency based response format was used for the Original questionnaire and Likert for the Other questionnaire. In Group 2 the frequency based response format was used for the Other questionnaire and the Likert for the Original questionnaire.

[a] $p<.05$ [b] $p<.01$ [c] $p<.001$;