

		Volume 52, issue 5, April 2012	ISSN 0191-8869
<b>PERSONALITY AND INDIVIDUAL DIFFERENCES</b>			
AN INTERNATIONAL JOURNAL OF RESEARCH INTO THE STRUCTURE AND DEVELOPMENT OF PERSONALITY, AND THE CAUSATION OF INDIVIDUAL DIFFERENCES			
Editors-in-Chief Dr T. VERNON, Canada Dr S. B. G. EYSENCK, London		Founding Editor Professor H. J. EYSENCK*	
<b>Contents:</b>			
<b>General Articles</b>		563	The International Society for the Study of Individual Differences (ISSID)
G.L. Flett, D.S. Molnar, T. Nepon and P.L. Hewitt		565	A mediational model of perfectionistic automatic thoughts and psychosomatic symptoms: The roles of negative affect and daily hassles
H.M. Baughman, S. Dearing, E. Giammarco and P.A. Vernon		571	Relationships between bullying behaviours and the Dark Triad: A study with adults
D.J. Kruger and J.S. Pigiowski		576	The effect of eyelid constriction on perceptions of mating strategy: Beware of the squinty-eyed guy!
E. Meins, S. McCarthy-Jones, C. Fernyhough, G. Lewis, R.P. Bentall and L.B. Alloy		581	Assessing negative cognitive style: Development and validation of a Short-Form version of the Cognitive Style Questionnaire
T. Podubinski, M. Daffern and S. Lee		586	A prospective examination of the stability of hostile-dominance and its relationship to paranoia over a one-year follow-up
H. Murtrie, J.S. Baxter, M.C. Obonsawin and S.C. Hunter		591	The relationship between memory beliefs, compliance and response change within a simulated forensic interview
K. Lee and M.C. Ashton		596	Getting mad and getting even: Agreeableness and Honesty-Humility as predictors of revenge intentions
M.M. McDonald, M.B. Donnellan and C.D. Navarrete		601	A life history approach to understanding the Dark Triad
P.K. Jonason and D.M. Buss		606	Avoiding entangling commitments: Tactics for implementing a short-term mating strategy
<i>[Continued on outside back cover]</i>			
<small>Person. Individ. Diff. is indexed/abstracted in: ASSIA, Curr. Cont. Soc. &amp; Behav. Sci., PASCAL-CNRS Data, Psychol. Abstr., PsycINFO, PsycLIT, Res. Alert, Soc. Sci. Cit. Indx. Also covered in the abstract and citation database SciVerse SCOPUS®. Full text available on SciVerse ScienceDirect®.</small>			
		<b>ISSN 0191-8869</b> 52(5) 363-656 (2012)	
OFFICIAL JOURNAL OF THE INTERNATIONAL SOCIETY FOR THE STUDY OF INDIVIDUAL DIFFERENCES (ISSID)			

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at SciVerse ScienceDirect

# Personality and Individual Differences

journal homepage: [www.elsevier.com/locate/paid](http://www.elsevier.com/locate/paid)

## The relationship of scale reliability and validity to respondent inconsistency

Michael D. Biderman<sup>a</sup>, Craig M. Reddock<sup>b,\*</sup>

<sup>a</sup> University of Tennessee at Chattanooga, Department of Psychology/2803, 350F Holt Hall, 615 McCallie Avenue, Chattanooga, TN 37403, United States

<sup>b</sup> Old Dominion University, 250 Mills Godwin Building, Department of Psychology, Norfolk, VA 23529, United States

### ARTICLE INFO

#### Article history:

Received 5 August 2011

Received in revised form 30 November 2011

Accepted 11 December 2011

Available online 11 January 2012

#### Keywords:

Intra-individual variability

Personality

Reliability

Validity

Inconsistency

Traitedness

### ABSTRACT

The generalizability of respondent inconsistency and of its relationship to estimates of the reliability and validity of personality measures was examined. Two hundred and six undergraduates completed two IPIP, the Thompson Mini-Markers measures of the Big Five and three other questionnaires. Respondent inconsistency was calculated based on average within-dimension standard deviations from one of the IPIP questionnaires and used to form three groups of respondents differing in mean inconsistency on that questionnaire. When groups formed from the first questionnaire were compared, mean reliabilities of the 13 other scales were largest when estimated using the most consistent respondents. Convergent validities between the other two scales of all Big Five measures were larger, and validity of conscientiousness computed from those scales as a predictor of GPA was higher in groups composed of more consistent respondents. Implications for consideration of inconsistency as a stable personality characteristic are considered.

© 2011 Elsevier Ltd. All rights reserved.

### 1. Introduction

When one thinks of reliability, probably the first thing that comes to mind is the test that is being considered. Reliability estimates are typically reported as characteristics of tests or scales. Rarely are the characteristics of the respondents, the test takers, mentioned when reliability is reported. However, when sources of variance in responses to test items are listed (e.g., Schmidt, Le, & Ilies, 2003), the errors that suppress reliability are attributed to the respondent or to the interaction between the respondent and the content of the test items. For example, Schmidt et al. listed momentary variations in attention, mental efficiency, and distractions on a given occasion – all respondent characteristics – as causes of random response error, one of the types of error that lower test reliability. They listed respondent-specific interpretation of the wording of questionnaire items, an interaction between respondents and items, as a source of specific factor error, another type of error that lowers test reliability. Although characteristics of the test items can affect these errors, in either case, the characteristics of the respondents clearly may affect the estimates of test or scale reliability as well.

Respondent characteristics also are rarely mentioned in descriptions of how to maximize reliability. If respondent characteristics are mentioned at all, the suggestion that reliability estimates depend on sample heterogeneity may be given, and most texts devoted to scale

development mention only the representativeness of pilot samples for the population in which the scale will be used. One characteristic that has been given little attention until recently is respondent inconsistency – the tendency for respondents to give different responses to items for which identical or nearly identical responses would seem to be appropriate. Such response inconsistency may be due to item wording differences. However, it may also be the case that respondent characteristics lead to inconsistent responses to theoretically equivalent items – items representing the same dimension on a personality inventory, for example.

Within-person inconsistency has been studied from several points of view, including the study of metatraits (Britt, 1993), or traitedness (Britt, 1993; Dwight, Wolf, & Golden, 2002), extreme response style (e.g., Greenleaf, 1992), and variability of test scores across time (e.g., Eid & Diener, 1999; Kernis, 2005). In recent studies some evidence that within-person variability may be viewed as a separate personality characteristic has been presented (Baird, Kimdy, & Lucas, 2006; Edwards & Woehr, 2007; Fleeson, 2001; Fleisher & Woehr, 2008). These studies suggested that within-person variability was a stable characteristic across time and that it moderated personality–performance relationships. Recently, Reddock, Biderman, and Nguyen (2011) found that within-person variability, defined based on responses to a Big Five personality questionnaire, exhibited incremental validity over both conscientiousness and cognitive ability in prediction of an academic performance criterion. These investigators found that respondents exhibiting larger within-person variability in responses to the questionnaire had smaller estimates of reliability of scales from that questionnaire than those who exhibited smaller within-person variability.

\* Corresponding author. Tel.: +1 757 683 6052.

E-mail addresses: [Michael-Biderman@utc.edu](mailto:Michael-Biderman@utc.edu) (M.D. Biderman), [credd005@old.edu](mailto:credd005@old.edu) (C.M. Reddock).

A limitation of their finding, however, was that the within-person variability was measured on the same scale on which differences in reliability were found. Thus, the relationship they found was one that might have been observed simply due to random differences in respondent inconsistency and misattributed as a stable characteristic. The present study extended this finding by measuring inconsistency using one questionnaire and then examining the relationship of inconsistency from that questionnaire to reliability and validity of other questionnaires. If respondent inconsistency were a random phenomenon, then it would not be expected to propagate across questionnaires, and inconsistent responders as measured on one questionnaire would not exhibit lower reliability on other questionnaires. On the other hand, if respondent inconsistency is a stable characteristic, then inconsistent responders identified on one questionnaire would exhibit lower reliability in different questionnaires.

Because of the relationship of reliability to validity, it would be expected that any factor that affects reliability would also affect validity. For that reason, it was expected that both convergent and criterion-related validity would be highest among respondents who exhibited the least inconsistency. To investigate this issue Big Five scale scores were computed from two questionnaires and the relationship of convergent validity of the Big Five scores between the two questionnaires to inconsistency defined by a separate questionnaire was assessed. To assess criterion-related validity, conscientiousness from each of the two questionnaires not used to define inconsistency was used to predict end-of-semester undergraduate grade point averages (GPA) of respondents. The relationship of the criterion-related validity of conscientiousness to respondent inconsistency defined by only the first questionnaire was assessed.

## 2. Method

### 2.1. Participants

Participants were 206 undergraduates at a medium sized southeastern University, participating for course credit. Sixty-four were male. Mean age was 19.32 ( $SD = 4.86$ ). Percentage of Whites was 67.96, Black/African-American was 25.24, and other was 6.80.

### 2.2. Measures

#### 2.2.1. Big Five Questionnaire 1

The scale used to measure respondent inconsistency was the 50-item Big Five Sample Scale from the web site of the International Personality Item Pool (IPIP; [www.ipip.ori.org](http://www.ipip.ori.org); Goldberg, 1999). In this questionnaire there are 10 items per Big Five dimension. Each item is a descriptive phrase, such as "I am the life of the party". It was administered via paper and pencil to all participants under instructions to respond honestly.

#### 2.2.2. Big Five Questionnaire 2

A second IPIP Big Five questionnaire was used in the assessment of reliability and validity. It consisted of the 50 items from the 100-item Sample Scale that were not in the 50-item Sample Scale.

#### 2.2.3. Thompson Mini-Markers

The Thompson Mini-Markers (Thompson, 2008) represents an alternative measure of the Big Five. This questionnaire consists of 40 items, with eight items per Big five dimension. Each item is a single word rather than a descriptive phrase. Examples are "Shy" and "Talkative". Test-retest reliabilities of the scales ranged from .70 to .83 in Thompson's (2008) development sample. Convergent validities with Goldberg's (1992) bipolar five-factor inventory

ranged from .68 to .78. On all questionnaires, respondents indicated how accurately each word described them on a 7-point response scale with 1 indicating "Completely Inaccurate" to 7 indicating "Completely Accurate". Responses to items on the following personality questionnaires were also on 7-point scales.

#### 2.2.4. Other scales

The Wonderlic Personnel Test (Wonderlic, 1999) and three other questionnaires were administered as part of a separate investigation relating a common method factor to self-report of affect (Biderman, Nguyen, & Cunningham, 2011). These were the Costello and Comrey (1967) depression scale, the Rosenberg self esteem scale (Rosenberg, 1965) and a shortened version of the Bateman and Crant (1993) proactive personality scale. The depression scale consists of 14 items, such as "I feel blue and depressed." The self-esteem scale consists of 10 items, such as "I feel I have a number of good qualities." The third scale consisted of 14 items, such as "I am constantly on the lookout for new ways to improve my life." For the purposes of this investigation these scales are viewed as a sample of types of scales that might be administered in a psychological evaluation and thus the relationship of their reliabilities to inconsistency are investigated here.

#### 2.2.5. GPA

Undergraduate grade point averages were obtained from school records immediately after completion of the semester in which respondents filled out the questionnaire. GPA values ranged from 0.115 to 3.085.

#### 2.2.6. Within-person Inconsistency (V)

Within-person inconsistency was measured in two ways. First, for each participant the standard deviation of that participant's responses to the 10 items from each Big Five dimension in Big Five Scale 1 was calculated, and the mean of the five standard deviations was computed. This measure is labeled unadjusted inconsistency. One potential problem with this measure is that it may be confounded with central tendency. That is, individuals who used only the low or only the high end of the response options may be labeled as consistent because of floor or ceiling effects. Although we used a 7-point response scale to overcome this limitation, we also computed a second measure of inconsistency, called adjusted inconsistency (Baird et al., 2006). For this measure the standard deviation of responses to items for each Big Five dimension was regressed onto the scale score and the square of the scale score for that dimension. The purpose of the regression analysis was to account for the relationship of variability of responses to response level. Thus, differences in variability due to differences in response level would be represented by predicted values in the regression analysis. Variability of the residuals from the regression would represent variability that was unrelated to the level of responses. The mean of the residuals from these five analyses was computed and represented inconsistency adjusted for response level.

#### 2.2.7. Reliability

Reliability was estimated using Cronbach's coefficient alpha. That coefficient was chosen because of its widespread use as an estimate of reliability (DeVellis, 2012). Although the assumptions for use of alpha as an estimate of reliability may not have been met for all of the scales used here, it is our belief that the results will generalize to other estimates of reliability.

### 2.3. Procedure

Participants completed the paper and pencil questionnaires in small groups. After participants filled out an informed consent sheet, the WPT was administered, followed by the 100-item IPIP

Big Five scale (from which the Big Five Scales 1 and 2 were taken), the Thompson Mini-Marker questionnaire, the depression, self-esteem and the proactive personality questionnaires in that order.

### 3. Results

Table 1 presents summary statistics of the scales administered to respondents and of the standard deviations of items within dimensions. Means and standard deviations of the two measures of inconsistency are also presented. Inspection of the table shows that for many of the dimensions the minimum value of the standard deviation of responses was zero, indicating that one or more respondents gave identical responses to all items within the dimension. For the IPIP questionnaires this was the case for no more than three respondents in any scale. For the Mini-Markers, depression, self-esteem, and proactive personality scales the numbers of respondents giving identical responses to all items were 12, 28, 5, and 1, respectively. Table 2 presents correlations of the domain scores with each other and with the two inconsistency measures and correlations of the scale standard deviations with each other and with the inconsistency measures. Inspection of Table 2 reveals that the standard deviations are fairly highly intercorrelated, and that the measures of inconsistency exhibit low correlations with the scale scores but substantial correlations with the scale standard deviations. The only exceptions are the relatively low correlations of the inconsistency measures with the standard deviations of the depression and self-esteem measures. At the present time we have no explanation for these low correlations.

The positive correlations between the individual dimension standard deviations lead to reliability coefficients of .602 for the unadjusted inconsistency measure and .700 for the adjusted measure. Although the adjusted measure of inconsistency was presented here to provide an indication of its relationship to the unadjusted measure, the results that follow are based on only the unadjusted measure. Corresponding results based on the adjusted measure were not different enough to warrant separate reports.

To examine the effect of inconsistency on reliability estimates, three groups of respondents were created based on the Questionnaire 1 unadjusted inconsistency – those least inconsistent, those with a midrange of inconsistency, and those most inconsistent. The three-way split of the distribution was done so that there were 68 or 69 persons at each level of inconsistency. Means values of

unadjusted inconsistency for each group were .90, 1.16, and 1.43 for the low, medium and high inconsistency groups, respectively.

Reliability estimates for all the study variables for each inconsistency group are presented in Table 3. In the top panel are reliabilities from the scales used to compute the inconsistency measure. Because coefficient alpha is based on the relationship of common variance to total variance (DeVellis, 2012) and because variance that is common to all items, as represented by interitem correlation coefficients, for example, decreases as response inconsistency increases, it would be expected that there would be differences in reliability between the three groups for Big Five Questionnaire 1 – the questionnaire used to define inconsistency – even if the differences in inconsistency were random across respondents. As expected the mean of reliability estimates from the least inconsistent group to the most inconsistent group were .887, .841, and .762.

A more appropriate comparison is of reliabilities of the scales not used to measure inconsistency. The bottom panel of Table 3 presents the reliability estimates from the scales whose responses were not used to define inconsistency. Means of the reliability estimates ranged from .864 for the group least inconsistent when responding to Big Five Scale 1 down to .780 for the most inconsistent group. A one way analysis of variance was performed on reliability estimates in the bottom panel of Table 3. The null hypothesis of equal mean reliabilities across inconsistency groups was rejected ( $F(2,36) = 5.00, p < .05$ ) with  $R^2$  equal .218. Post-hoc tests using Tukey's b statistic indicated that the mean of the most inconsistent group was significantly different from the mean of the least inconsistent group.

To assess the extent to which convergent validity was related to inconsistency group, correlations between corresponding Big Five dimensions in the two questionnaires not used to define inconsistency were computed. Those correlations are presented in Table 4. Mean convergent validity across the five dimensions between the two questionnaires decreased as average inconsistency defined using Questionnaire 1 increased. The statistical significance of differences across groups was assessed using multigroup correlation analyses in Amos (Arbuckle, 1983/2010). Specifically a three-group model was created for Questionnaire 2 and Mini-Marker domain scores in which covariances between all domain scores in the two questionnaires were estimated separately for each group. This was a completely saturated model with degrees of freedom equal to 0. Then a second model was applied with covariances between

**Table 1**  
Summary statistics of domain scores and of the standard deviations of items within dimensions.

	Domain scores				Standard deviations			
	Mean	SD	Min	Max	Mean	SD	Min	Max
1 E	4.75	1.04	2.10	7.00	1.14	0.41	0.00	2.27
1 A	5.30	0.74	2.60	7.00	1.04	0.39	0.00	2.47
1 C	4.58	0.86	1.90	7.00	1.15	0.36	0.00	2.15
1 S	4.24	1.00	2.20	6.60	1.30	0.41	0.47	2.53
1 O	4.85	0.82	1.80	6.80	1.20	0.40	0.32	2.39
2 E	4.82	0.86	2.10	7.00	1.20	0.44	0.00	2.75
2 A	5.05	0.72	2.30	6.80	1.30	0.46	0.42	2.57
2 C	4.79	0.95	1.80	7.00	1.21	0.43	0.00	2.41
2 S	4.22	0.93	1.80	6.60	1.29	0.43	0.32	2.42
2 O	4.76	0.85	2.10	7.00	1.33	0.51	0.00	2.85
M E	4.98	1.03	2.25	7.00	1.03	0.49	0.00	2.85
M A	5.66	0.78	1.50	7.00	0.89	0.50	0.00	2.78
M C	4.90	1.01	1.75	7.00	0.99	0.49	0.00	2.78
M S	3.88	0.99	1.50	6.63	1.23	0.52	0.00	3.02
M O	5.08	0.96	1.00	7.00	1.15	0.54	0.00	2.66
Depression	1.84	0.83	1.00	6.71	0.78	0.48	0.00	2.32
Self-esteem	5.65	0.87	1.90	7.00	1.12	0.52	0.00	2.46
Proactive personality	4.85	0.98	2.07	7.00	1.41	0.57	0.00	2.84
Unadjusted inconsistency	1.16	0.24	0.53	2.11				
Adjusted inconsistency	0.00	0.26	-0.71	0.89				

**Table 2**  
Correlations among domain scores, standard deviations and inconsistency measures. Correlations between domain scores are below the diagonal, those between within-scale standard deviations are above the diagonal. Correlations involving inconsistency measures appear both below and above the diagonal. Whole sample reliability coefficients are on the diagonal.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1 Q1 E	<b>.895</b>	.227 <sup>c</sup>	.158 <sup>a</sup>	.285 <sup>c</sup>	.142 <sup>a</sup>	.418 <sup>c</sup>	.353 <sup>c</sup>	.099	.271 <sup>c</sup>	.241 <sup>c</sup>	.386 <sup>c</sup>	.017	-.005	.127	.171 <sup>a</sup>	-.013	.040	.081	.595 <sup>c</sup>	.557 <sup>c</sup>
2 Q1 A	.317 <sup>c</sup>	<b>.789</b>	.140 <sup>a</sup>	.336 <sup>c</sup>	.231 <sup>c</sup>	.098	.208 <sup>b</sup>	.058	.189	.234 <sup>c</sup>	.120	.117	.152 <sup>a</sup>	.080	.048	.149 <sup>a</sup>	.050	.021	.621 <sup>c</sup>	.563 <sup>c</sup>
3 Q1 C	.007	.164 <sup>b</sup>	<b>.823</b>	.217 <sup>b</sup>	.313 <sup>c</sup>	.222 <sup>c</sup>	.146 <sup>a</sup>	.476 <sup>c</sup>	.215 <sup>b</sup>	.266 <sup>c</sup>	.237 <sup>c</sup>	.083	.374 <sup>c</sup>	.240 <sup>c</sup>	.165 <sup>a</sup>	.037	.078	.209 <sup>b</sup>	.567 <sup>c</sup>	.559 <sup>c</sup>
4 Q1 S	.237 <sup>b</sup>	.176 <sup>a</sup>	-.021	<b>.842</b>	.265 <sup>c</sup>	.310 <sup>c</sup>	.370 <sup>c</sup>	.204 <sup>b</sup>	.451 <sup>c</sup>	.206 <sup>b</sup>	.205 <sup>b</sup>	.198 <sup>b</sup>	.125	.348 <sup>c</sup>	.274 <sup>c</sup>	.054	.065	.167 <sup>a</sup>	.688 <sup>c</sup>	.684 <sup>c</sup>
5 Q1 O	.244 <sup>c</sup>	.335 <sup>c</sup>	.270 <sup>c</sup>	.156 <sup>a</sup>	<b>.812</b>	.367 <sup>c</sup>	.344 <sup>c</sup>	.293 <sup>c</sup>	.270 <sup>c</sup>	.546 <sup>c</sup>	.158 <sup>a</sup>	.228 <sup>c</sup>	.253 <sup>c</sup>	.207 <sup>b</sup>	.390 <sup>c</sup>	.089	.153 <sup>a</sup>	.183 <sup>b</sup>	.630 <sup>c</sup>	.621 <sup>c</sup>
6 Q2 E	.813 <sup>c</sup>	.377 <sup>c</sup>	.186 <sup>b</sup>	.286 <sup>c</sup>	.462 <sup>c</sup>	<b>.800</b>	.460 <sup>c</sup>	.286 <sup>c</sup>	.422 <sup>c</sup>	.478 <sup>c</sup>	.348 <sup>c</sup>	.157 <sup>a</sup>	.189 <sup>b</sup>	.288 <sup>c</sup>	.306 <sup>c</sup>	-.060	.063	.106	.461 <sup>c</sup>	.498 <sup>c</sup>
7 Q2 A	.423 <sup>c</sup>	.786 <sup>c</sup>	.237 <sup>b</sup>	.286 <sup>c</sup>	.378 <sup>c</sup>	.499 <sup>c</sup>	<b>.704</b>	.288 <sup>c</sup>	.406 <sup>c</sup>	.363 <sup>c</sup>	.271 <sup>c</sup>	.234 <sup>c</sup>	.196 <sup>b</sup>	.221 <sup>c</sup>	.330 <sup>c</sup>	.056	.216 <sup>b</sup>	.253 <sup>c</sup>	.464 <sup>c</sup>	.484 <sup>c</sup>
8 Q2 C	.058	.239 <sup>b</sup>	.848 <sup>c</sup>	-.039	.250 <sup>c</sup>	.219 <sup>b</sup>	.324 <sup>c</sup>	<b>.834</b>	.258 <sup>c</sup>	.367 <sup>c</sup>	.155 <sup>a</sup>	.073	.544 <sup>c</sup>	.221 <sup>c</sup>	.280 <sup>c</sup>	-.018	.147 <sup>a</sup>	.219 <sup>b</sup>	.357 <sup>c</sup>	.395 <sup>c</sup>
9 Q2 S	.112	.107	.032	.827 <sup>c</sup>	.127	.156 <sup>a</sup>	.179 <sup>b</sup>	.007	<b>.801</b>	.309 <sup>c</sup>	.263 <sup>c</sup>	.164 <sup>a</sup>	.154 <sup>a</sup>	.284 <sup>c</sup>	.290 <sup>c</sup>	.034	.105	.095	.453 <sup>c</sup>	.516 <sup>c</sup>
10 Q2 O	.318 <sup>c</sup>	.317 <sup>c</sup>	.265 <sup>c</sup>	.296 <sup>c</sup>	.758 <sup>c</sup>	.505 <sup>c</sup>	.399 <sup>c</sup>	.295 <sup>c</sup>	.241 <sup>c</sup>	<b>.769</b>	.227 <sup>c</sup>	.153 <sup>a</sup>	.291 <sup>c</sup>	.119	.302 <sup>c</sup>	.131	.158 <sup>a</sup>	.107	.482 <sup>c</sup>	.475 <sup>c</sup>
11 M-M E	.838 <sup>c</sup>	.308 <sup>c</sup>	.026	.170 <sup>a</sup>	.221 <sup>b</sup>	.782 <sup>c</sup>	.401 <sup>c</sup>	.093	.063	.276 <sup>c</sup>	<b>.871</b>	.107	.173 <sup>a</sup>	.164 <sup>a</sup>	.180 <sup>b</sup>	.063	.061	.118	.347 <sup>c</sup>	.345 <sup>c</sup>
12 M-M A	.159 <sup>a</sup>	.631 <sup>c</sup>	.249 <sup>c</sup>	.187 <sup>b</sup>	.190 <sup>b</sup>	.237 <sup>b</sup>	.640 <sup>c</sup>	.378 <sup>c</sup>	.141 <sup>a</sup>	.197 <sup>b</sup>	.193 <sup>b</sup>	<b>.811</b>	.083	.122	.263 <sup>c</sup>	.254 <sup>c</sup>	.131	.099	.208 <sup>b</sup>	.220 <sup>c</sup>
13 M-M C	-.005	.115	.830 <sup>c</sup>	-.044	.231 <sup>c</sup>	.143 <sup>a</sup>	.206 <sup>b</sup>	.830 <sup>c</sup>	.009	.264 <sup>b</sup>	.047	.306 <sup>c</sup>	<b>.865</b>	.323 <sup>c</sup>	.182 <sup>b</sup>	.168 <sup>a</sup>	.142 <sup>a</sup>	.232 <sup>c</sup>	.281 <sup>c</sup>	.293 <sup>c</sup>
14 M-M S	.157 <sup>a</sup>	.005	.007	.721 <sup>c</sup>	.165 <sup>a</sup>	.212 <sup>b</sup>	.179 <sup>b</sup>	-.049	.711 <sup>c</sup>	.220 <sup>b</sup>	.115	.111	-.047	<b>.804</b>	.170 <sup>a</sup>	.131	.080	.234 <sup>c</sup>	.324 <sup>c</sup>	.371 <sup>c</sup>
15 M-M O	.177 <sup>b</sup>	.216 <sup>b</sup>	.198 <sup>b</sup>	.175 <sup>a</sup>	.762 <sup>c</sup>	.405 <sup>c</sup>	.275 <sup>c</sup>	.268 <sup>c</sup>	.118	.677 <sup>c</sup>	.189 <sup>b</sup>	.226 <sup>b</sup>	.233 <sup>b</sup>	.157 <sup>a</sup>	<b>.825</b>	.045	.148 <sup>a</sup>	.196 <sup>b</sup>	.341 <sup>c</sup>	.348 <sup>c</sup>
16 Dep	-.202 <sup>b</sup>	-.309 <sup>c</sup>	-.330 <sup>c</sup>	-.284 <sup>c</sup>	-.192 <sup>b</sup>	-.311 <sup>c</sup>	-.449 <sup>c</sup>	-.340 <sup>c</sup>	-.181 <sup>b</sup>	-.295 <sup>c</sup>	-.267 <sup>c</sup>	-.400 <sup>c</sup>	-.302 <sup>c</sup>	-.220 <sup>b</sup>	-.200 <sup>b</sup>	<b>.920</b>	.244 <sup>c</sup>	.014	.101	.027
17 RSE	.285 <sup>c</sup>	.188 <sup>b</sup>	.381 <sup>c</sup>	.242 <sup>c</sup>	.359 <sup>c</sup>	.407 <sup>c</sup>	.386 <sup>c</sup>	.414 <sup>c</sup>	.156 <sup>a</sup>	.404 <sup>c</sup>	.269 <sup>c</sup>	.264 <sup>c</sup>	.389 <sup>c</sup>	.273 <sup>c</sup>	.423 <sup>c</sup>	-.674 <sup>c</sup>	<b>.847</b>	.207 <sup>b</sup>	.125	.117
18 PP	.247 <sup>c</sup>	.176 <sup>a</sup>	.338 <sup>c</sup>	.134	.453 <sup>c</sup>	.389 <sup>c</sup>	.318 <sup>c</sup>	.385 <sup>c</sup>	.042	.467 <sup>c</sup>	.235 <sup>b</sup>	.212 <sup>b</sup>	.349 <sup>a</sup>	.081	.469 <sup>c</sup>	-.416 <sup>c</sup>	.586 <sup>c</sup>	<b>.874</b>	.211 <sup>b</sup>	.252 <sup>c</sup>
19 Unadj V	.002	-.005	.036	-.174 <sup>a</sup>	.022	-.011	.019	.175 <sup>a</sup>	-.162 <sup>a</sup>	-.009	.080	.199 <sup>b</sup>	.195 <sup>b</sup>	-.276 <sup>c</sup>	.080	-.029	.043	.009	<b>.602</b>	.962 <sup>c</sup>
20 Adj V	.162 <sup>a</sup>	.161 <sup>a</sup>	.096	-.056	.152 <sup>a</sup>	.158 <sup>a</sup>	.184 <sup>b</sup>	.224 <sup>b</sup>	-.075	.126	.213 <sup>b</sup>	.288 <sup>c</sup>	.228 <sup>b</sup>	-.198 <sup>b</sup>	.156 <sup>a</sup>	-.101	.111	.079	.962 <sup>c</sup>	<b>.700</b>

<sup>a</sup>  $p < .05$ .  
<sup>b</sup>  $p < .01$ .  
<sup>c</sup>  $p < .001$ .

**Table 3**  
Reliability estimates for each group defined by the unadjusted inconsistency measure computed from Big Five Scale 1.

Scale	Inconsistency		
	Low	Medium	High
Big Five Questionnaire 1 E	.920	.904	.843
Big Five Questionnaire 1 A	.898	.783	.676
Bid Five Questionnaire 1 C	.824	.849	.799
Big Five Questionnaire 1 S	.900	.841	.789
Big Five Questionnaire 1 O	.897	.832	.705
Mean reliability of Questionnaire 1 scales	.887	.841	.762
SDs of reliabilities of Questionnaire 1 scales	.037	.043	.069
Big Five Questionnaire 2 E	.842	.824	.743
Big Five Questionnaire 2 A	.826	.687	.562
Big Five Questionnaire 2 C	.784	.841	.840
Big Five Questionnaire 2 S	.813	.803	.785
Big Five Questionnaire 2 O	.867	.781	.663
Mini-Marker E	.907	.882	.825
Mini-Marker A	.902	.728	.763
Mini-Marker C	.868	.872	.849
Mini-Marker S	.809	.812	.772
Mini-Marker O	.876	.835	.774
Depression	.951	.884	.908
Self esteem	.869	.839	.839
Proactive Personality	.924	.853	.832
Mean reliability of other scales	.864	.819	.780
SDs of reliabilities of other scales	.048	.058	.090

**Table 4**  
Convergent validities between Questionnaire 2 and Mini-Marker Big Five domain scores across inconsistency groups.

Scale	Inconsistency group		
	Low	Medium	High
Questionnaire 2 E with Mini-Marker E	.882	.767	.714
Questionnaire 2 A with Mini-Marker A	.802	.559	.507
Questionnaire 2 C with Mini-Marker C	.827	.830	.818
Questionnaire 2 S with Mini-Marker S	.731	.726	.660
Questionnaire 2 O with Mini-Marker O	.690	.662	.692
Mean convergent validity	.786	.708	.678
SDs of convergent validities	.076	.104	.103

corresponding dimensions in the two questionnaires, e.g., Questionnaire 2 E with Mini-Marker E, constrained to be equal across

**Table 5**  
Criterion-related validities of three measures of conscientiousness in prediction of GPA for the three inconsistency groups.

Scale	Inconsistency Group			
	Overall	Low	Medium	High
Big Five Questionnaire 2 C	.191	.185	.375	.078
Mini-Marker C	.194	.263	.346	.050

the three groups. This constraint created a special model with 10 degrees of freedom. The chi-square difference between the two models was 20.30 ( $p < .05$ ), suggesting that the covariances between corresponding domain scores were related to inconsistency.

Criterion-related validity defined as correlations of GPA with measures of conscientiousness from Big Five Questionnaire 2 and the Mini-Marker questionnaire were computed for each group. Table 5 presents the validity coefficients. Inspection of the table shows that validity was roughly the same for both the most consistent group and the middle group but fell off dramatically for the most inconsistent group for both scales. To provide some evidence of the significance of differences across inconsistency groups, two multigroup regression models were formed using Amos. In the first, GPA was regressed simultaneously onto Questionnaire 2 and Mini-Marker Conscientiousness domain scores allowing regression coefficients and variances to be unique within each inconsistency group, creating a saturated model. In the second, restricted model, variances and regression weights were constrained to be equal across groups, yielding eight degrees of freedom. The chi-square difference statistic was 18.10 ( $p < .05$ ), suggesting that criterion-related validity was related to inconsistency.

#### 4. Discussion

This study was conducted to examine the generalizability of respondent inconsistency and of its relationship to reliability, convergent and criterion-related validity of commonly used personality questionnaires. This research replicated and extended key previous findings concerning the stability of within-person variability.

First, we found that within a Big Five instrument, inconsistency of responding to each of the dimensions was stable enough to allow an overall measure of inconsistency to have marginally acceptable reliability. Second, an overall measure of inconsistency computed from the responses to Questionnaire 1 was related to the mean of reliability estimates from the 13 other scales in the study. Mean reliability of most inconsistent respondents was about .08 less than mean reliability of the most consistent respondents. This is evidence for the stability of the inconsistency measure across at least the duration it took to administer the several questionnaires used in this research. Since that relationship was significant with a moderate effect size, the results provide evidence consistent with the suggestion that within-person variability is a general personality characteristic related to estimates of reliability across personality scales. These results extend those of Reddock et al. (2011) and agree with those of Fleeson (2001), who measured inconsistency across periods of many days and found persons more inconsistent at one time to be more inconsistent at other times.

Finally, the results provide further evidence for the relationship of within-person inconsistency to both convergent validity and criterion-related validity of personality tests. In both the analyses of convergent validity and of criterion-related validity, means of validity estimates were lowest for the most inconsistent group. For example, with respect to criterion-related validity, if only medium or low inconsistent respondents were considered, validity could be .3 or larger. On the other hand, if the sample included a large number of inconsistent respondents, validity could be below .1.

The possibility that inconsistency is a result of difficulty in processing negatively-worded items was raised by an anonymous reviewer. To assess this possibility, estimates of unadjusted inconsistency for the positively worded items within each dimension and for the negatively-worded items were computed for Questionnaire 1. The correlation between the two measures was 0.53. The mean values of unadjusted inconsistency for the two sets of items were 1.06 and 1.12 with standard deviations 0.27 and 0.32 for positively- and negatively-worded items, respectively. Although the difference in means was significant ( $t = 3.21, p < .01$ ) the proportion of variance due to item wording was quite small, indicating that a majority of the variance in inconsistency was due to individual differences in respondents.

The extent to which results for inconsistency as defined here were related extreme response style (Greenleaf, 1992) was assessed by computing a measure of extreme response style as the proportion of extreme responses (1s and 7s on the response scale) for Questionnaire 1 (e.g., see Naemi, Beal, & Payne, 2009). There was a positive correlation between the measure of extreme response style and the unadjusted inconsistency ( $r = .424, p < .001$ ). To insure that the differences found above were not due to differences in extreme response style, three extreme response style groups were formed in the same fashion as inconsistency groups, and reliabilities of the 13 scales not part of Questionnaire 1 were obtained. Mean reliabilities of the 13 scales across the three groups from the group with least extreme response style to the most extreme were .78, .80, and .82, a trend opposite the trend for inconsistency, and not significantly different ( $p > .05$ ). Mean convergent validities between Questionnaire 2 and Mini-Marker scale scores across the three extreme response style groups were .64, .68, and .76, significantly different ( $p < .01$ ) but again with the trend opposite that for the three inconsistency groups. Finally, criterion-related validity decreased across the three extreme response style

groups, a trend similar to that for the inconsistency groups. These comparisons suggest that reliability and validity of scales may also be related to extreme response style although the relationships are not identical to and may be more complicated than the relationships to inconsistency of responding.

This study adds to a growing body of research indicating that inconsistency is a personal characteristic that may play an important role in the evaluation of personality scales, particularly for employee selection. Future research to assess the temporal stability of the measure of inconsistency used here is needed. A more detailed investigation of the joint relationships of inconsistency and extreme response style to reliability and validity is also called for.

## References

- Arbuckle, J. L. (1983/2010). *Amos 19.0.0*. Meadville, PA: Amos Development Corporation.
- Baird, B. M., Kimdy, L., & 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*, 512–527.
- Bateman, T. S., & Crant, J. M. (1993). The proactive component of organizational behavior: A measure and correlates. *Journal of Organizational Behavior, 14*, 103–118.
- Biderman, M. D., Nguyen, N. T., & Cunningham, C. J. L. (2011). A method factor measure of self-concept. Paper presented at the 26th Annual Conference of The Society for Industrial and Organizational Psychology, Chicago, IL.
- Britt, T. (1993). Metraits: Evidence relevant to the validity of the construct and its implications. *Journal of Personality and Social Psychology, 65*, 554–562.
- Costello, C. G., & Comrey, A. L. (1967). Scales for measuring depression and anxiety. *The Journal of Psychology, 66*, 303–313.
- DeVellis, R. F. (2012). *Scale development: Theory and applications*. Thousand Oaks, CA: Sage.
- Dwight, S. A., Wolf, P. P., & Golden, J. H. (2002). Metraits: Enhancing criterion-related validity through the assessment of traitedness. *Journal of Applied Social Psychology, 32*, 2202–2212.
- Edwards, B. D., & Woehr, D. J. (2007). An examination and evaluation of frequency-based personality measurement. *Personality and Individual Differences, 43*(4), 803–814.
- Eid, M., & Diener, E. (1999). Intra-individual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology, 76*, 662–676.
- 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. (2008). The big six? The importance of within-person personality consistency in predicting performance. Paper presented at the meeting of the Southern Management Association, St. Petersburg, FL.
- Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Journal of Personality and Social Psychology, 59*, 1216–1229.
- 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.
- Greenleaf, E. A. (1992). Measuring extreme response style. *The Public Opinion Quarterly, 56*, 328–351.
- Kernis, M. H. (2005). Measuring self-esteem in context: The importance of stability of self-esteem in psychological functioning. *Journal of Personality, 73*, 1569–1605.
- Naemi, B. D., Beal, D. J., & Payne, S. C. (2009). Personality predictors of extreme response style. *Journal of Personality, 77*, 261–286.
- Reddock, C. M., Biderman, M. D., & Nguyen, N. T. (2011). The relationship of reliability and validity of personality tests to frame-of-reference instructions and within-person inconsistency. *International Journal of Selection and Assessment, 19*(2), 119–131.
- Rosenberg, M. (1965). *Society and the adolescent self image*. Princeton, NJ: Princeton University Press.
- Schmidt, F. L., Le, H., & Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual differences constructs. *Psychological Methods, 8*, 206–224.
- Thompson, E. R. (2008). Development and validation of an international English Big-Five Mini-Markers. *Personality and Individual Differences, 45*, 542–548.
- Wonderlic Inc. (1999). *Wonderlic's personnel test manual and scoring guide*. Chicago, IL: Author.