

Maximizing validity of personality questionnaires

Michael D. Biderman

River Cities Industrial-Organizational
Psychology Conference
University of Tennessee at Chattanooga
2014

Thanks to

Nhung Nguyen
Towson University

for her work in collecting and understanding
the data of sample 3.

The problem

It's generally believed and much evidence supports that **cognitive ability** is the best single predictor of performance in a variety of settings.

The best personality predictor is generally thought to be **conscientiousness**.

Alas, the validity of conscientiousness is about 34% as large as validity of general mental ability (Schmidt, 2012).

Conventional wisdom re personality questionnaires is summarized by the following quote, "The problem with personality tests is . . . that the validity of personality measures as predictors of job performance is often disappointingly low." Morgeson et al. (2007).

This presentation examines three procedures that might increase validity of personality measures.

The focus of the presentation is on Big Five personality questionnaires – those that measure extraversion, agreeableness, conscientiousness, stability, and openness.

The three procedures presented here are alternative ways of analyzing data already collected – ways that may increase the validity of predictions.

The data . . .

A four-sample combined dataset. N=1143.

Biderman, M. D., Worthy, R., & Nguyen, N. T. (2012). N = 328.

Reddock, C. M., Biderman, M. D., & Nguyen, N. T. (2010). N = 310.

Nguyen, N. T., & Biderman, M. D. (2013). N = 299.

Biderman, M. D., Nguyen, N. T., & Cunningham, C. J. L. (2011). N = 206.

All participants were given IPIP 50-item Sample Big Five questionnaire (www.ipip.ori.org)

All given Wonderlic Personnel Test (WPT) – a measure of cognitive ability.

End of semester GPA recorded from academic records for all participants.

Basic Results . . .

Validity of WPT for the 4-sample data set: **0.272.**

Validity of Conscientiousness: **.180.** (66% of the WPT r)

As we would expect, cognitive ability was more valid as a predictor of GPA than was conscientiousness.

Why not use ACT or SAT validity as the goal?

ACT, while influenced by cognitive ability, is also a measure of other qualities.

We're interested in comparing personality to cognitive ability only.

For that reason, we'll focus on ways to make personality as valid as the WPT in this presentation.

Procedure Number 1: Use **all** the scales measured by the questionnaire.

Big Five questionnaires yield 5 measures. Why focus on only conscientiousness? Use them all.

Form a **multiple** regression equation. Include all the domains that the questionnaire was designed to measure.

For the Big Five, this means perform a simultaneous multiple regression of GPA onto ALL FIVE Big Five scales.

The Multiple regression result for the 4-sample data:

Recall: Validity of WPT = **.272** <--- Our goal.

Validity of all 5 Big Five domain scores in a simultaneous regression . . .

Multiple R = .254.

Since we're ganging up on WPT by using multiple predictors, from now on, only Rs adjusted for the number of predictors will be reported.

Adjusted Multiple R = **.245**

So we've gained on WPT without gathering any more data from our participants, just by using the responses sitting in our computer.

We also may have learned something about GPA and personality. GPA exhibited significant relationships to FOUR of the 5 scale scores in the multiple regression.

<u>Scale Score</u>	<u>Standardized Coef</u>	<u>p</u>
Extraversion	-.120	.000
Agreeableness	.080	.010
Conscientiousness	.184	.000
Stability	-.109	.000
Openness	-.012	.709

Procedure Number 2: Use Latent Variable Analyses

Latent variable techniques are touted as a way to estimate relationships correcting for biases caused by measurement error (Schmidt et al., 2013).

Measurement error contaminates scale scores, but the effect of that contamination is removed when relationships are estimated using latent variable analyses.

We have evidence of another source of contamination.

We believe that Big Five item responses are very likely contaminated by personal characteristics of the respondents other than those ostensibly measured by the Big Five (Biderman et al., 2011).

Those other personal characteristics are easily measured and their effects removed by latent variable analyses.

We have estimated three contaminating characteristics from Big Five questionnaires –

1. The general affective state of the respondent.
2. A tendency to agree with positively-worded items.
3. A tendency to agree with negatively-worded items.

What could be the effects of these characteristics?

Two possibilities

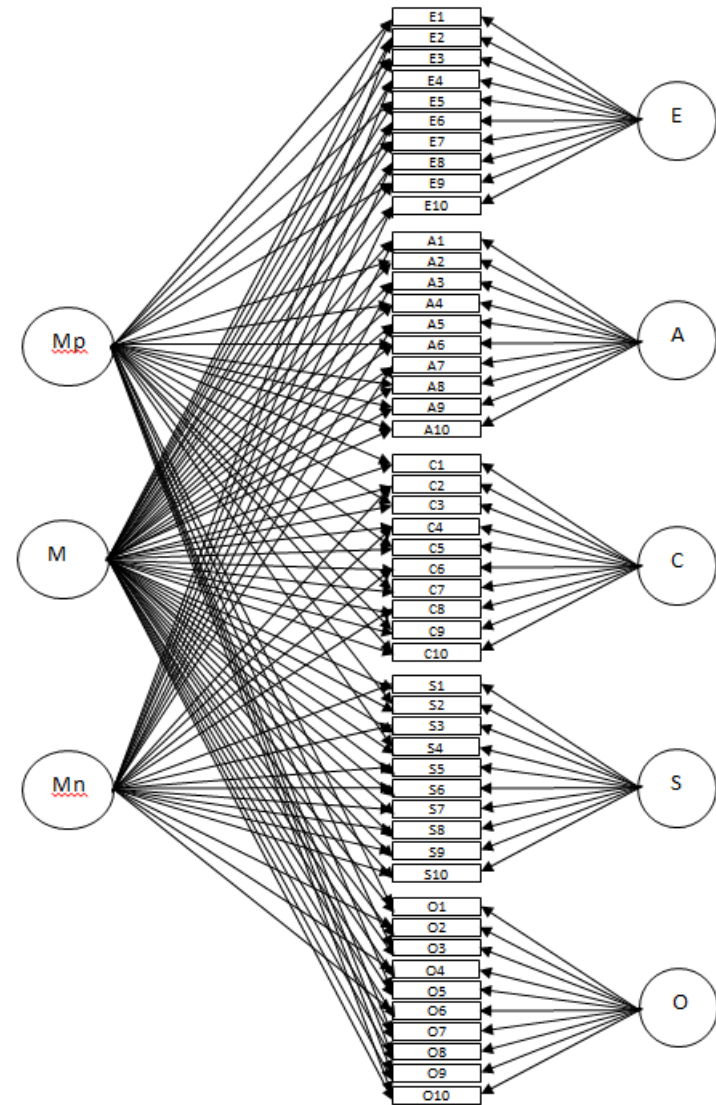
1. They could act just as error variance acts - as noise - reducing the estimated strengths of relationships of Big Five scales scores to criteria.

2. The tendencies may themselves be valid predictors of criteria. They may themselves predict the criteria.

Whatever is the case, if these tendencies exist, individual differences in them should be separated from individual differences in the Big Five domains in any validity assessment.

A Path Diagram of the Latent Variable Model showing how the three contaminants affect Big Five items

Those familiar with such path diagrams will note that all factors in this model are orthogonal.



Are these tendencies present in our data?

Goodness of fit measurements . . .

<u>Model</u>	<u>Chi-square</u>	<u>df</u>	<u>CFI</u>	<u>RMSEA</u>	<u>SRMR</u>	<u>Chi-sq Diff</u>	<u>df</u>
Full Model	4276.165	1075	.857	.051	.061		
Sans M	5454.814	1125	.807	.058	.086	1178.649	50
Sans Mp	4987.878	1101	.827	.055	.065	711.713	26
Sans Mn	5018.657	1099	.825	.055	.066	742.492	26

The above results show that all three added factors contribute significantly to goodness-of-fit – so they’re all valuable additions to the model.

These results suggest that the three new tendencies ARE present in and affecting our data.

Validity of the Latent Variable Model for the 4-sample data

The latent variable model was applied using Mplus.

Factor scores were computed by Mplus for each Big Five domain factors and for the three new factors.

The factor scores were put into the SPSS data editor.

GPA values were regressed onto the EIGHT factor scores.

Factor Score Validity results . . .

Adjusted Multiple R = **.316**. (Validity of WPT = .272)

<u>Factor Score</u>	<u>Standardized Coef</u>	<u>p</u>
Extraversion	-.070	.000
Agreeableness	.071	.015
Conscientiousness	.084	.004
Stability	-.146	.000
Openness	-.088	.002
M (Exp of affectivity)	-.164	.000
Mp (Pos item bias)	.130	.000
Mn (Neg item bias)	.109	.000

Procedure Number 3: Using response inconsistency

In previous work, (Reddock & Biderman, 2011) we found evidence suggesting that persons differ in the inconsistency with which they respond to personality questionnaires.

Some persons give the same response to each item from a specific domain. Others with the same level of the trait give high and low responses that average out to that level.

I decided to investigate inconsistency of responses as a predictor.

Domain Inconsistency = Standard deviation of responses to items from the same domain.

Overall inconsistency = mean of 5 Domain Inconsistency values.
Alpha = .65

I added Overall Inconsistency to the 8 factor scores.

Multiple regression of GPA onto 8 factor scores + Inconsistency

Adjusted Multiple R = **.338**.

<u>Factor Score</u>	<u>Standardized Coef</u>	<u>p</u>
Extraversion	-.086	.003
Agreeableness	.059	.041
Conscientiousness	.073	.010
Stability	-.169	.000
Openness	-.050	.095
M (Exp of affectivity)	-.138	.000
Mp (Pos item bias)	.149	.000
Mn (Neg item bias)	.101	.001
Inconsistency	-.133	.000

Summary of Results

1. The validity of a single personality questionnaire – the IPIP 50-item Sample Questionnaire – was almost doubled by re-analyzing the data using recently developed techniques.
2. For these data, the validity was at least as large as that of a well-regarded cognitive ability test.
3. All but one of the factors were incrementally valid.
4. None of what was done involved gathering more data.

Limitations

1. Only the **IPIP** 50-item Sample Questionnaire was used. The results must be replicated for other questionnaires – the NEO, HEXACO, etc.
2. The regression results were optimal for this sample. We would expect shrinkage on cross-validation.
3. Validity, though as large as that for the WPT, was not large. Only about 10% of variance in GPA was accounted for.
4. The latent variable approach advocated here, while not unfamiliar to psychologists, is definitely not yet mainstream.

Conclusion

As stated in title to the keynote address,

Personality at Work: It is More than Conscientiousness

We can add,

Personality at School: It is More than Conscientiousness.

Questions?

References

- Biderman, M. D., Nguyen, N. T. & Cunningham, C. J. L., & Ghorbani, N. (2011). The ubiquity of common method variance: The case of the Big Five. *Journal of Research in Personality, 45*, 417-429.
- 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, April.
- Biderman, M. D., Worthy, R., & Nguyen, N. T. (2012). Manipulating method variance. Paper presented at the 27th annual conference of The Society for Industrial and Organizational Psychology, San Diego, CA, April.
- Nguyen, N. T., & Biderman, M. D. (2013). Predicting counterproductive work behavior from a bifactor model of Big Five personality. Paper presented at the meeting of the Academy of Management, Orlando, FL, August.
- Reddock, C. M., Biderman, M. D., & Nguyen, N. T. (2010). Increasing the validity of personality questionnaires. Paper presented at the 25th annual conference of The Society for Industrial and Organizational Psychology, Atlanta, GA, April.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007) Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60*, 683-729.
- Schmidt, F. L. (2012). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 100 years of research. Keynote presented at the 8th annual River Cities Industrial-Organizational Psychology conference, Chattanooga, TN.
- Schmidt, F. L., Le, H., & Oh, I. (2013). Are true scores and construct scores the same? A critical examination of their substitutability and the implications for research results. *International Journal of Selection and Assessment, 21*, 339-354.

Logistic Regression – 8 Factor scores + Inconsistency

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	efsmmpmnob5	-.120	.070	2.959	1	.085	.887
	afsmmpmnob5	.138	.073	3.598	1	.058	1.148
	cfsmmpmnob5	.146	.072	4.086	1	.043	1.157
	sfsmpmnob5	-.445	.077	33.707	1	.000	.641
	ofsmmpmnob5	.230	.077	8.867	1	.003	1.259
	mfsmpmnob5	-.233	.073	10.279	1	.001	.792
	mpfsmpmnob5	.223	.080	7.818	1	.005	1.249
	mnfsmpmnob5	.261	.075	12.024	1	.001	1.298
	meansd	-.818	.263	9.695	1	.002	.441
Constant	.909	.298	9.279	1	.002	2.482	

a. Variable(s) entered on step 1: efsmmpmnob5, afsmmpmnob5, cfsmmpmnob5, sfsmpmnob5, ofsmmpmnob5, mfsmpmnob5, mpfsmpmnob5, mnfsmpmnob5, meansd.

Inconsistency as a moderator of the C -> GPA relationship – Linear Regression

Coefficients ^a				Standardized		
Model	Unstandardized Coefficients		Std. Error	Coefficients	t	Sig.
	B			Beta		
1	(Constant)	3.327	.086		38.612	.000
	efsmmpmnob5	-.061	.020	-.086	-2.993	.003
	afsmmpmnob5	.040	.021	.054	1.869	.062
	cfsmpmnob5	.322	.096	.435	3.349	.001
	sfsmpmnob5	-.122	.022	-.167	-5.562	.000
	ofsmmpmnob5	.035	.022	.046	1.548	.122
	mfsmpmnob5	-.101	.021	-.138	-4.793	.000
	mpfsmpmnob5	.108	.023	.137	4.668	.000
	mnfsmpmnob5	.073	.022	.098	3.382	.001
	meansd	-.310	.075	-.127	-4.121	.000
	cfsXmeansd	-.223	.078	-.369	-2.854	.004

a. Dependent Variable: GPA

$$\hat{Y} = ? + .322C - .310*Inc - .223*C*Inc = ?? - .310*Inc + (.322 - .223Inc)*C$$

The larger the value of Inc, the more shallow the slope of C.

So the more Inconsistent the participant, the weaker the C->GPA relationship.

C does not predict well for inconsistent responders.

Inconsistency as a moderator of the C -> GPA relationship – Logistic Regression

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	efsmmpmnob5	-.118	.070	2.858	1	.091	.888
	afsmmpmnob5	.131	.073	3.208	1	.073	1.140
	cfsmpmnob5	.859	.341	6.333	1	.012	2.360
	sfsmpmnob5	-.444	.077	33.370	1	.000	.642
	ofsmmpmnob5	.224	.078	8.368	1	.004	1.252
	mfsmpmnob5	-.233	.073	10.285	1	.001	.792
	mpfsmpmnob5	.197	.080	6.037	1	.014	1.218
	mnfsmpmnob5	.257	.075	11.651	1	.001	1.293
	meansd	-.804	.265	9.223	1	.002	.447
	cfsXmeansd	-.599	.280	4.583	1	.032	.549
	Constant	.887	.301	8.683	1	.003	2.429

a. Variable(s) entered on step 1: efsmmpmnob5, afsmmpmnob5, cfsmpmnob5, sfsmpmnob5, ofsmmpmnob5, mfsmpmnob5, mpfsmpmnob5, mnfsmpmnob5, meansd, cfsXmeansd.