Validity of a Contextualized Conscientiousness Scale for Ph.D. Candidate Performance

René Butter[1] & Marise Born[2]

Abstract

The present study examined the criterion validity of a contextualized conscientiousness scale developed specifically for the prediction of the performance of Ph.D. candidates. This scale more strongly predicted relevant Ph.D. performance criteria than global Big Five measures did, which supports the development of customized tests.

Press paragraph

Successful performance of Ph.D. candidates is of direct importance to the success of academia. Little is known yet about the relation between Ph.D. personality and performance. A scale was custom-developed for Ph.D. candidates to measure conscientiousness, which is among the best personality predictors of performance. The custom-developed scale significantly related to relevant performance criteria, such as progress of the research and the estimated probability to complete the project in time. The scale had incremental validity over the more general Big Five personality dimensions, implying that a customized approach may contribute to improving the criterion validity of personality measures in general.

---

[1] René Butter Psychologisch Advies (RBPA)
[2] Erasmas University Rotterdam; the authors thank Thomas Faith, Miram Heemskerk and Jojanneke Struijs for their help in collecting the data

Validity of a Contextualized Conscientiousness Scale for Ph.D. Candidate Performance

A consistent finding across several meta-analyses is that conscientiousness proves to be the best personality predictor of performance criteria. Moreover, the added value of the other four personality dimensions above conscientiousness in the prediction of overall job performance is very low (cf. Morgeson et al., in press).

At the same time, however, the old predictive validity challenge for personality inventories still exists (e.g. Morgeson et al., 2007; Morgeson et al., in press), as they have relatively low levels of validity in comparison to other instruments. Therefore, several suggestions have been made to achieve better validity. Recently Dudley, Orvis, Lebiecki, and Cortina (2006) conducted a meta-analysis of the role of conscientiousness in the prediction of job performance. Confirming earlier meta-analyses, conscientiousness turned out to be the personality trait with the highest criterion validity. The authors improved predictive validity by using sub facets of the broad conscientiousness trait, resulting in incremental validity over the sole use of these broad traits. Their findings also confirmed the importance of a multidimensional perspective of job performance with different personality variables being more relevant in the prediction of different performance criteria. Nevertheless, narrow traits may still suffer from lack of face validity and ecological validity, because they are defined across situations and are not tailored to reflect a specific context.

Other authors focus on the frame-of-reference from which test takers complete personality test items. Each person may complete the same personality scale from a different perspective, for example from a private person perspective or from a work

perspective. These different perspectives will increase the error variance of the scales and hence reduce their criterion validity. To tackle this problem, Schmit, Ryan, Stierwalt, and Powell (1995) and Bing, Whanger, Davison, and VanHook (2004) added a specific context to each item. They used an altered conscientiousness scale on which a reference to school was appended to each item, usually at the beginning or the end of the statement. For example, the item "I strive for excellence in everything I do" was modified to "I strive for excellence in everything I do at school". They demonstrated that such items, labeled 'context-specific items', can lead to incremental criterion validity for conscientiousness as a predictor of job performance.

In our view adding such general context-information is useful to decrease the error variance in the measurements and thereby increase the criterion-validity. This approach may remain artificial, however, in the sense that general context information is added to standard test items regardless of whether these standard items were relevant for the specific context in the first place. It is unclear, for example, why a supposedly context-specific item such as "I always keep my things in a fixed place at work", is a relevant predictor for conscientiousness in research settings.

Other scholars take an even more critical approach towards the use of broad standard personality scales for predicting job performance than the frame-of-reference researchers. For example, in Morgeson et al. (2007, p. 715), Neal Schmitt claims that "he would avoid published personality measures in almost all instances and would construct his own measures that are linked directly to job tasks in a face valid or relevant fashion". This strategy of a customized development of tests is propagated as a way to increase

acceptance of personality tests, as they may be perceived as more authentic and reflective of the realities of test-takers and client organizations (e.g., Morgeson et al., in press).

The present study focuses on such a custom-developed test strategy, but not only for reasons of acceptability. Also it is expected that better predictive validity levels may be reached, using the same argument as by Schmit et al. (1995), namely measurement error reduction.

The conscientiousness scale in this study was developed for the position of Ph.D. candidate in The Netherlands. A recent study for the Dutch government (Berger & De Jonge, 2005), reported that the output rates in terms of number of Ph.D's obtained at the Dutch universities needs much improvement. Only four to eight percent of the Ph.D. candidates finish their Ph.D. within the generally prescribed four years. Between 17 and 33% need five years and about 10% leave university without a Ph.D. degree. It will be clear that especially this 10% drop-out leads to waste of human capital that is tragic for the person involved, involves a financial loss for the universities and a general socio-economic loss. It should be noted that unlike in the United States, most Ph.D. candidates in The Netherlands work in full-time post-Master research jobs at a university. Most of them apply for a paid job in which they conduct a study to obtain the Ph.D. degree. A more or less formal selection procedure mostly is utilized. The conclusions of Berger and De Jonge (2005) indicated that personality attributes are perceived to be one of the factors that determine the success of Ph.D. candidates. They furthermore concluded that the quality of coaching by the advisors of Ph.D. candidates is the most important factor considered to explain the output rates.

By improving our understanding of the relationship between personality characteristics and the performance of Ph.D. candidates we also aim to facilitate a more personalized coaching style by Ph.D. advisors.

A recent publication by Noordam and Gosling (2006) and their matching columns in Science (e.g. Noordam & Gosling, 2007) show that also internationally the attention has increased for non-academic aspects of the performance of Ph.D. candidates. However, these authors focus more on practical skills such as how to relate to colleagues and how to celebrate your success, than on the personality attributes underlying these skills. The international I/O-Psychological literature nevertheless hardly contains any references to the relation of personality factors and success of Ph.D. candidates. The closest related topic to our knowledge concerns the relation between personality factors and academic success on Bachelor's and Master's levels. In studies focusing on these levels, conscientiousness is consistently reported to be a positive predictor of academic success (Chamorro-Premuzic & Furnham, 2003, a; Chamorro-Premuzic & Furnham, 2003, b; Diseth, 2003; De Raad & Schouwenburg, 1996; Dollinger & Orf, 1991; Paunonen & Ashton, 2001).

As no specific conscientiousness scale was available for the level of Ph.D. candidates, we constructed our own scale to measure personality characteristics of these candidates. In line with the earlier mentioned recommendations in the recent literature, we decided to develop relevant, face-valid scales. Because of the relatively 'unexplored' nature of our object, we started off by using a qualitative research approach to obtain an understanding of the perspective of those involved in the 'natural surroundings' of Ph.D. candidates. Based on in-depth interviews with these stakeholders, rich contextual material

was collected on the basis of which we developed face valid test items. Unlike the general school reference that Schmit et al. (1995) added to each general personality item, we constructed items that each were fully derived from the specific context, in this case the 'natural world of Ph.D. candidates'. Using an environmental parallel, such items may be labeled 'ecological'. These items cover more traits, but in this paper we will concentrate on the items developed to measure conscientiousness.

Based on the finding reported in the meta-analysis by Dudley et al. (2006) we expect that our custom-made conscientiousness scale will correlate highest with performance measures seen as directly relevant for Ph.D. candidates (in this case with work progress and meeting deadlines: see method section). We furthermore expect that this scale will provide incremental validity above Big Five measures since its closer proximity to the work content will lead to greater face validity and therefore to less error variance.

Method

Construction of the customized conscientiousness scale

Qualitative interviews on success and failure factors (e.g., Cassell & Symon, 2004) were conducted with eight representatives of the relevant stakeholders. These included Ph.D. candidates, Ph.D. graduates, drop-outs, and Ph.D. advisors. Based on this information, scale items were constructed. We constructed the items in English because many Ph.D. candidates in The Netherlands originate from other countries and are therefore not fluent in Dutch. Moreover, also Dutch candidates need to master the English language. Finally, candidates working in foreign universities also participated in the study. Sixty-two items, 10 of which measure conscientiousness, were constructed and

completed by 242 Ph.D. candidates in an internet survey. Factor analysis (Principal Axis Factoring) followed by Varimax rotation yielded an interpretable orthogonal five factor solution that reflects the Big Five dimensions, indicating that the customized items behave like personality scale items. With respect to conscientiousness, we found a set of seven items loading on one factor with a reasonable internal consistency (a = .73). As can be seen in the measures subsection this reliability coefficient is somewhat lower than that of the Big Five conscientiousness scale. The content of the seven customized items reflects *accurate time management, working in a structured and self-propelled way, setting one's own research priorities and keeping appointments.*

Further evidence for the construct validity of the scale is provide by Table 1 that gives the correlations of the contextualized conscientiousness scale with the Big Five scales. The only significant correlations found with Big Five dimensions were for conscientiousness (r=.53, p<.01) and openness (r=.-.20, p<.05).

Participants and Procedure

The criterion related validity study was a web-based field survey in which the test-takers completed the contextualized scale and the Big Five scales in one session. The customized items were presented to each respondent in a random sequence to prevent order effects. These items were completed by 242 PhD. candidates from a wide array of academic disciplines varying from Psychology, Engineering, Medicine, Art Studies to Philosophy. The sample consists of participants of various nationalities from both Dutch and non-Dutch universities (less than 5%). As indicated, all questions and items were phrased in English, and no serious language problems were reported. Big Five scores are

available for 103 participants and performance ratings for 107 participants, partly

overlapping the 103 participants. The reason for this incomplete design is that not all

participants allowed us to contact their advisors. Also, the Big Five items were not

included in the earliest period of data collection.

In the first data collection step, we approached participants through their

institutions or through Ph.D. candidate networks. Next, we contacted them individually

using standardized mail and telephone protocols. We also asked the participants in the

web-based survey for their permission to contact their Ph.D. advisors in order to collect

performance ratings. For the advisors we developed a concise email questionnaire with

performance rating items. The advisors who were indicated by their Ph.D. candidates

were informed that a specific candidate had participated in our study and who had given

permission to contact the advisor in question. The advisor then was asked to complete the

performance rating items.

Not surprisingly, the data collection phase for this study was demanding. At

times, Ph.D. candidates were reluctant to give their permission. This is understandable

given the power distance and the potentially sensitive relationship between a Ph.D.

candidate and his or her advisor. An in our opinion moving example sheds some light on

aspects of this issue. One candidate kindly informed us that his situation was special in

view of an ADHD disorder and advised us to tag his responses in the database, thinking

these responses might distort the general picture in the sample. At the same time he asked

us to not inform his supervisor of his medical problem

Thus, many potential participants had to be contacted in order to obtain sufficient

participants. There are reasons to believe that well-achieving Ph.D. candidates are

overrepresented in the sample, potentially leading to reduced variance. Also, of the 107

Ph.D. candidates for whom we were able to collect performance ratings, about 25% were

rated by the same advisors. The overlap, however, is limited as only one advisor rated

four candidates and two advisors rated three candidates. Therefore, we did not take these

dependencies into further account in the data analysis for this paper.

Measures

To examine the incremental validity of the customized conscientiousness scale we

used a 50 item Big Five instrument taken from the International Personality Item Pool

(Goldberg, 1999). Each dimension of this instrument is measured by 10 items using a five

point Likert scale ranging from 1 (very inaccurate) to 5 (very accurate) for each of the

five dimensions. For example (for conscientiousness): "I follow a schedule".

Three of the fifty items were replaced by alternative ITIP items (*I get caught up in my*

*problem*s instead of *I often feel blue*) or slightly changed (*Neglect my duties* instead of

Shirk my duties and *Get tasks done right away* in stead of *Get chores done right away*),

because of potential language difficulties for non-native English speakers.

The scale reliabilities for the IPIP scales in our sample are, respectively, .90 for

extraversion, .77 for agreeableness, .80 for conscientiousness, .84 for emotional stability

and .74 for intellect. These are very similar to those reported by Goldberg (1999).

Criterion measures

Recent literature (Dudley et al., 2006, p. 43) states that "job performance is

multidimensional and that different personality variables may be more relevant in the

prediction of difference performance criteria ". In line with this assumption, we used a

wide set of performance criteria covering various aspects of the Ph.D. work. The

following 9 dimensions were measured using 10-point Likert scales ranging from (1 very

poor) to 10 (excellent): 1) general functioning ($M = 7,.88$, $SD = 0.97$), 2) academic

quality ($M = 7.87$, $SD = 0.78$), 3) work progress ($M = 7.49$, $SD = 1.13$), 4) performance

with respect to publications ($M = 7.30$, $SD = 1.09$), 5) teaching ($M = 7.81$, $SD = 0.88$), 6)

meeting deadlines ($M = 7.52$, $SD = 1.51$), 7) performance during conferences ($M = 7.85$,

$SD = 0.92$), 8) interpersonal functioning at the university ($M = 8.12$, $SD = 0.92$), 9)

estimated probability to obtain the Ph.D. within the prescribed period (mostly four years)

(M= 81%, SD = 18%). We obtained criterion ratings for 107 candidates. As was

indicated above, the relation between a Ph.D. candidate and the supervisor is sensitive.

As can be seen above, the variance in the criterion measures is not very high, which

possibly is caused by an overrepresentation of well-achieving participants.

Results

Before testing the hypotheses, we first inspected the correlations of the customized conscientiousness scale with the performance criteria. Table 1 reports these results. Significant correlations were found with work progress ($r = .25, p < .05$), meeting deadlines ($r = .33, p < .01$) and the estimated probability to obtain the Ph.D. within the prescribed period ($r = .40, p < .01$).

Table 2 gives the multiple regression analysis results for each of these performance criteria separately. Only relevant performance criteria (work progress, meeting deadlines and the estimated probability) could be predicted by the customized conscientiousness scale, whereas the other contextualized scales (that is for openness, extraversion, agreeableness and emotional stability) were not predictive of these criteria. This discriminant pattern can be regarded as further support for the construct validity of the customized conscientiousness scale. The regression analyses of the performance criteria on the Big Five scales yielded no significant results at all. Apparently, these scales that are devoid of context, showed no predictive power.

To investigate the incremental validity of the customized conscientiousness scale above that of the big five measures, hierarchical regression analyses were performed.

Table 3 gives the results, showing that the customized conscientiousness scale has a significant incremental predictive power above the Big Five dimensions for these three criteria. That is, for work progress, meeting deadlines and the estimated probability to obtain the Ph.D. within the prescribed period.

In sum, the findings were the following:

1) The customized conscientiousness scale correlates highest with the Big Five

conscientiousness scale

2) It is significantly related to directly relevant performance criteria but not related to the

other, less relevant performance criteria that we used for Ph.D. candidates

3) In predicting these performance criteria the customized scale shows incremental

validity above the Big Five scales

4) This incremental criterion validity was achieved at relatively high measurement

efficiency, that is, 7 items vs. 10 per Big Five scale.

5) The incremental criterion validity was realized despite a lower reliability than that of

the Big Five scales (a = .73 vs. .80 for the corresponding Big Five scale).

Thus, we may conclude that the customized conscientiousness scale has an adequate

construct validity and shows incremental validity above the Big Five.

Discussion

The theoretical issue at stake here is whether customized or ecological items have incremental criterion validity over that explained by general personality factors. Or, in other words, do we believe that behavior is stable across situations or that the interaction between person and context forms the heart and soul of face-valid and predictive scales? Our analyses seem to more strongly support the latter point of view. Customized, or ecological conscientiousness items give better prediction of job performance of Ph.D. candidates than general Big Five factors do.

The practical aim of this study was to shed more light on the relation between personality factors and performance of Ph.D. candidates. Until now, not much is known on this topic. Based on our study we can conclude that (customized) conscientiousness plays an important role here. We think that it is useful to take into account other personality characteristics as well, when allocating Ph.D. candidates to projects. In the Political Sciences domain, for example, Ph.D. projects may contain a large philosophical component, but other projects may imply people skills like participating in governmental project teams and interviewing politicians. Based on our interviews, there is reason to believe that the competencies needed for specific projects are not always examined when allocating Ph.D. candidates to projects. Also, opinion surveys among Ph.D. candidates indicate that they experience a high need for more personalized coaching (Berger & De Jonge, 2005; Meijer, 2002). This suggest that lack of coaching may be a factor that contributes to the high drop-out rates, at least in The Netherlands, and to corresponding personal, social and economic losses. We are afraid that the situation in other countries will be rather similar.

The aim of this study was to have a better understanding of the relation between personality and performance of Ph.D. candidates. This insight is essential to improve personal coaching, selection and competency management in general for Ph.D. candidates.

As the present study dealt with a very specific context, we were highly doubtful whether off-the-shelf general personality instruments provided by the testing industry could be meaningfully applied here. For example, one can be a very sloppy person with an untidy room, but still be a highly systemized researcher. The literature to some extent supports our primal thought that if the relation between personality and performance of Ph.D. candidates is to be explored in a fruitful way, it will have to be done with a specifically developed instrument. As we started off with qualitative interviews, we constructed contextualized items in a bottom-up, data-driven way. Based on recent literature, we also took into account a wide range of nine performance criteria, because there are strong indications that personality factors correlate higher with directly relevant criteria.

We focused on conscientiousness because this construct is the most consistent predictor of job performance over a wide range of jobs. Our practical conclusion is that the customized conscientiousness scale has a reasonable reliability as well as construct validity and is related to relevant performance criteria, that is, work progress, meeting deadlines and the estimated probability to complete the Ph.D. project in time.

Our study had its limitations. Firstly, cognitive ability was not included in the design. The reason for this first of all was that we were primarily interested in the relation between personality and performance. Also, Ph.D. candidates are a ve ry selective group

in the cognitive sense. Therefore, we had doubts whether cognitive factors would have predictive power here. Also, given our web-based field survey it was not possible to guarantee the testing circumstances needed to complete cognitive tests. Finally, adding cognitive tests most likely would have further lowered the response rate. A second limitation is that the sample to investigate incremental validity is not large (N=60). All other analyses are based on larger samples, however.

Through this study, we hope to have supported our belief that the road to more valid personality tests is paved by developing items that are intrinsically relevant for specific contexts. We believe that customized items can contribute to this goal.

References

Berger, J, & de Jonge, J. (2005). *Rendement verkend. Succes en faalfactoren van promotietrajecten aan Nederlandse Universiteiten* [Output explored. Success and failure factors in Ph.D. trajectories at Dutch Universities]. Zoetermeer, Netherlands: EIM Onderzoek voor Bedrijf en Beleid.

Bing, M. N., Whanger, J. C., Davison, H. K., & VanHook, J. B. (2004). Incremental validity of the frame-of-reference effect in personality scale scores: a replication and extension. *Journal of Applied Psychology, 89*, 150-157.

Cassell, C., & Symon, G. (Eds.) (2004). *Essential guide to qualitative methods in organizational research. London*: Sage.

Chamorro-Premuzic, T., & Furnham, A. (2003a). Personality traits and academic examination performance. *European Journal of Personality, 17*, 237-250.

Chamorro-Premuzic, T., & Furnham, A. (2003b). Personality predicts academic performance: evidence from two longitudinal university samples. *Journal of Research in Personality*, 37, 319-338.traits and academic

De Raad, B., & Schouwenburg, H. C. (1996). Personality in learning and education. *European Journal of Personality, 10*, 303-336.

Diseth, A. (2003). Personality and approaches to learning as predictors of academic achievement. *European Journal of Personality, 17*, 143-155.

Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: examining the intercorrelations and the incremental validity of narrow traits.

Journal of Applied Psychology, *91*, 40-57.

Goldberg, L. R. (in press).A broad band‑with, public domain, personality inventory

    measuring the lower‑level facets of several five factor models. In I. Mervielde, I.

    Deary, F. De Fruyt, & F. Ostendorp (Eds), *Personality Psychology in Europe*

    (Vol. 7, pp. 7-28). Tilburg, Netherlands: Tilburg University Press.

Hough, L. M., & Oswald, F. L. (2005). They're right, well mostly right: research

    evidence and an agenda to rescue personality testing from the 1960s

    insights.*Human Performance*, *18*, 373-387.

Meijer, M. M. (2002). Behoud Talent!: *Een rapportage over de verschillende aspecten*

    die een rol spelen bij de begeleiding van promovendi. [Retain Talent!. A report

on several aspects playing a role in advising Ph.D. students]. Landelijk AIO en OIO

    overleg.

Morgeson, 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*, 683-72.

Morgeson, P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., &

    Schmitt, N. (in press), Are we getting fooled again?: Coming to terms with

    limitations in the use of personality tests for personnel solution. *Personnel*

    Psychology.

Noordam, B., & Gosling, P. (2006). *Mastering your Ph.D.: survival and success in the*

    doctoral years and beyond. New York: Springer.

Noordam, B., & Gosling, P. (2007). Mastering your Ph.D.: relating to your co‑workers'

personality types. Retrieved August, 31, 2007 from

http://sciencecareers.sciencemag.org/career_development/previous_issues/articles

/2007_07_20/caredit_a0700104/(parent)/68

Murphy, K.R., & Dzieweczynski, J.L. (2005). Why don't measures of broad dimensions

of personality perform better as predictors of job performance. *Human*

Performance, *18*, 343-357.

Paunonen, S. V., & Ashton (2001). Big Five Predictors of Academic Achievement.

Journal of Research in Personality, *35*, 78-90.

Schmit, J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference

effects on personality scale scores and criterion-related validity. *Journal of*

Applied Psychology, *80*, 607-620.

Table 1

Correlations of the customized conscientiousness scale with the Big Five dimensions (N=

103)

Big Five Dimension	Correlation
Extraversion	-.09
Agreeableness	.11
Openness	-.20*
Conscientiousness	.52**
Emotional stability	.14

* *p* < .05 (two-tailed)

** *p* < .01 (two-tailed)

Table 2

Regression of the performance criteria on customized conscientiousness (N=103).

Performance criteria	Beta for context specific conscientiousness	R
Work progress	.27**	.29**
Meeting deadlines	.34**	.38**
Estimated probability to obtain the Ph.D. within the prescribed period	.41**	.42**

Note. there are no other significant predictors. The regression results for the other six performance criteria were not significant.

** *p* < .001 (two-tailed)

Table 3

Hierarchical Regression Results of the performance criteria on the Big Five scales (first step) and customized conscientiousness (second step) (N=60)

Performance criteria	Beta	R^2	$? R^2$
Work progress	.38*	.20	.10*
Meeting deadlines	.35*	.17	.08*
Estimated probability to obtain the Ph.D. within the prescribed period	.40**	.20	.12**

* $p < .05$ (two-tailed)

** $p < .001$ (two-tailed)