

Supervisor ratings of students' academic potential as predictors of citizenship and counterproductive behavior

Citation for published version (APA):

Schwager, I. T. L., Hülshager, U. R., Lang, J., Klieger, D. M., Bridgeman, B., & Wendler, C. (2014). Supervisor ratings of students' academic potential as predictors of citizenship and counterproductive behavior. *Learning and Individual Differences*, 35, 62-69. <https://doi.org/10.1016/j.lindif.2014.07.005>

Document status and date:

Published: 01/01/2014

DOI:

[10.1016/j.lindif.2014.07.005](https://doi.org/10.1016/j.lindif.2014.07.005)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.



Supervisor ratings of students' academic potential as predictors of citizenship and counterproductive behavior



Inge T.L. Schwager^{a,*}, Ute R. Hülshager^a, Jonas W.B. Lang^b, David M. Klieger^c, Brent Bridgeman^c, Cathy Wendler^c

^a Department of Work and Social Psychology, Maastricht University, Maastricht, The Netherlands

^b Department of Personnel Management, Work and Organizational Psychology, Ghent University, Ghent, Belgium

^c Educational Testing Service, Princeton, NJ, United States

ARTICLE INFO

Article history:

Received 3 February 2014

Received in revised form 13 June 2014

Accepted 18 July 2014

Keywords:

Academic performance

University citizenship behavior

Counterproductive academic behavior

Noncognitive predictors

Diversity

ABSTRACT

The educational psychology literature has recently discussed the implications of university citizenship behavior – activities that go beyond the primary student role – and counterproductive academic behavior – behaviors that run counter to academic values and objectives. This study investigated whether supervisor ratings of students' academic potential can serve as a valuable predictor of those criteria. The sample included 115 international students enrolled in a master's degree program at a university in the Netherlands. Results revealed that supervisor ratings of students' academic potential provided incremental validity above and beyond undergraduate grade point average (U-GPA) in predicting self-ratings of university citizenship and counterproductive academic behavior. The usefulness of supervisor ratings as supplement to traditional predictors of study success, and implications for student selection are discussed.

© 2014 Elsevier Inc. All rights reserved.

Harvard University recently suspended more than 60 students due to a cheating scandal (Pérez-Peña, 2013). The suspended students collaborated on a take home exam, in spite of strict instructions to work on it autonomously. Their behavior clearly had a negative impact on the reputation of the elite university, which ironically uses the Latin word for truth as its official slogan. Apparently, not all students who were admitted were strongly committed to Harvard's integrity principle. At the graduate level, cheating is a widespread issue as well (McCabe, Butterfield, & Trevino, 2006). This is especially alarming as graduate students are about to enter the job market where the consequences of unethical behavior are yet more severe compared to those in the academic context.

In efforts to take additional aspects of students' performance into account, researchers have recently suggested conceptualizing academic performance in broader terms (Lievens, Buyse, & Sackett, 2005; Meriac, 2012; Schmitt, Oswald, Friede, Imus, & Merritt, 2008; Sinha, Oswald, Imus, & Schmitt, 2011; Zettler, 2011). It has initially been acknowledged in the field of personnel selection, that performance is multidimensional, subsuming not only task but also contextual performance aspects such as citizenship and counterproductive behavior (Rotundo & Sackett, 2002; Viswesvaran & Ones, 2000). Task performance

describes the fulfillment of tasks and responsibilities that are formally associated with a given job or role, whereas citizenship behavior refers to activities that are not necessarily related to the primary tasks of a given role, but that facilitate organizational effectiveness by contributing to the organizational, social and psychological environment (Borman & Motowidlo, 1997). While these two dimensions focus on activities that promote performance and benefit an organization, counterproductive behavior describes activities that stand counter to an organization's values and objectives and that harm the well-being of an organization (Rotundo & Sackett, 2002).

In a similar way, academic performance is multidimensional and entails not only task-related, but also university citizenship and counterproductive academic behavior (Meriac, 2012; Schmitt et al., 2008; Zettler, 2011). Task performance is certainly important in order to meet the formal requirements of a study program. University citizenship behavior, however, might also have a positive influence on students' academic success by stimulating cooperation between peer students, or faculty members, which in turn increases the likelihood of getting helpful guidance or feedback (Verduin & Clark, 1991). Counterproductive academic behavior, in contrast, interferes more directly with academic success, in cases where students do not attend a class or cheat on exams instead of learning the study materials (Credé & Niehorster, 2009). The major goals of universities are to prepare students for the job market (Morphew & Hartley, 2006), and to foster their employability (Poropat, 2011). Thus, the multidimensional conceptualization of performance plays an important role in the academic context as well.

* Corresponding author at: Department of Work and Social Psychology, Maastricht University, Universiteitssingel 40, 6229 ES Maastricht, The Netherlands. Tel.: +31 43 388 12 58.

E-mail address: inge.schwager@maastrichtuniversity.nl (I.T.L. Schwager).

This is best represented by universities' mission statements that often highlight the importance of interpersonal and psychological aspects of study success (Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004). Furthermore, a recent meta-analysis provided evidence for the positive influence of education on not only future task performance, but also future citizenship and counterproductive work behavior (Ng & Feldman, 2009).

Although task performance is best represented by students' grades on exams or assignments, examples of university citizenship behavior include helping fellow students understand the subject matter, accepting inconvenient study conditions or engaging in voluntary activities. Counterproductive academic behavior includes behaviors such as copying from published sources, cheating on exams, stealing university property or bullying individuals affiliated with the university. Students sharing their knowledge to help other students provide better support to universities' educational missions than students who try to improve their own performance by cheating or stealing study materials. Accordingly, universities might benefit from selecting students based on their tendencies to engage in university citizenship and productive academic behavior, to increase the possibility of approving students who support universities' well-functioning instead of negatively affecting their reputation. In sum, higher levels of university citizenship behavior and reduced levels of counterproductive academic behavior have a positive influence on both, not only the students, but also the universities that benefit from a counterbalanced and committed student body.

In order to find an appropriate way to predict university citizenship and counterproductive academic behavior, we broaden the predictor domain and investigate the predictive validity of supervisor ratings of students' academic potential. Based on the two assumptions that past behavior predicts future behavior (Wernimont & Campbell, 1968) and that supervisory ratings are a reliable predictor of typical performance (Marcus, Goffin, Johnston, & Rohstein, 2007), we hypothesize that ratings of students' noncognitive academic potential provided by former supervisors are a valid predictor of future university citizenship and counterproductive academic behavior. Specifically, we investigate the predictive validity of the ETS® Personal Potential Index (ETS® PPI), a web-based evaluation system that uses former supervisors' ratings to systematically assess noncognitive aspects of academic success. Given that the utility of a new selection approach is provided only in case additional variance is explained over existing and widely used selection instruments, we also investigate the incremental validity of the ETS PPI above and beyond undergraduate grade point average (U-GPA).

1. Noncognitive predictors of academic performance and study success

Several researchers have already defined study success in broader terms and identified factors and instruments that might relate to this expanded conceptualization. For instance, in an early study, Enright and Gitomer (1989) interviewed university faculty members and identified seven competencies that are important for graduate study success: communication, creativity, explanation, motivation, planning, professionalism, and synthesis. In order to measure the seven dimensions, Enright and Gitomer (1989) proposed to use simulation tasks but they also critically discussed that using such simulation tasks are likely to be difficult to implement in large-scale student selection.

More recently, Oswald et al. (2004) systematically analyzed universities' mission statements and educational objectives in order to identify dimensions of undergraduate academic performance in broad sense. They defined twelve major dimensions which can be sub-categorized as intellectual behaviors (knowledge, learning, artistic), interpersonal behaviors (multicultural, leadership, interpersonal, citizenship), and intrapersonal behaviors (health, career, adaptability, perseverance, ethics). Based on their findings they developed a situational judgment test (SJT) and a biographical data (biodata) questionnaire. SJTs can be used to assess noncognitive skills, including descriptions of problematic

situations individuals are instructed to solve. Answers are typically provided in a multiple response format or on a Likert scale. Biodata measures include structured questions about applicants' background and life history. SJTs and biodata measures have good predictive validities and even reduce subgroup differences in combination with cognitive predictors (Lievens & Coetsier, 2002; McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001; Oswald et al., 2004; Schmitt et al., 2009). However, the influence of coaching and faking is still an issue that needs to be taken into account when using these methods in practice (Ramsay, Schmitt, Oswald, Kim, & Gillespie, 2006).

Letters of recommendations are another option to assess a broader conceptualization of performance while reducing the effects of social desirable responding. Among universities they are widely used (Liu, Minsky, Ling, & Kyllonen, 2009). They are provided by former supervisors and include useful insights about students' knowledge, skills and personality (McCarthy & Goffin, 2001). Because they are based on other-ratings, they are less vulnerable to socially-desirable responding and cheating. A notable shortcoming of letters of recommendations is, however, that they suffer from typically weak psychometric properties (Range et al., 1991). As no guidelines exist for writing letters of recommendation, they are all written differently making it difficult to compare between applicants, and also to get a coherent picture of one applicant providing several letters. Standardization has been described as a successful means to improve the reliability and validity of letters of recommendation (Kim & Kyllonen, 2007; Kyllonen & Kim, 2005; McCarthy & Goffin, 2001; Nicklin & Roch, 2009; Schmitt, 2012) and faculty members and administrators prefer standardized letters of recommendation over traditional letters (Walters, Kyllonen, & Plante, 2006).

The ETS PPI is an instrument that combines advantages of traditional letters of recommendation (insights from former supervisors) with advantages of standardized rating forms (better psychometric properties). In contrast to traditional letters of recommendation which have no predetermined format or content, the ETS PPI measures students' academic potential with a set of items targeting six dimensions, thereby assessing the same content across applicants in a standardized way. Importantly, the ETS PPI relies on ratings from multiple (up to 5) supervisors, thereby enhancing reliability and validity. Previous research has documented that ratings from single other-raters are, to some extent, idiosyncratic which reduces reliability and attenuates correlations with outcome variables (Connelly & Ones, 2010). Using a composite score that is based on multiple raters is a way to overcome rater idiosyncrasies and increase reliability and validity (Connelly & Ones, 2010).

The six ETS PPI dimensions (see Table 1) are the result of interviews with deans and faculty members working at numerous universities (Briel et al., 2000), and a more in-depth follow-up interview phase (Walpole, Burton, Kanyi, & Jackenthal, 2001). Walpole et al. (2001) asked faculty members to identify variables they felt were predictive of graduate school success and that could be used as part of the admissions process. A number of noncognitive variables were ranked highly (e.g., persistence, tenacity, collegiality, communication, and enthusiasm) and functioned as basis for the ETS PPI. Except for the teamwork dimension, all ETS PPI dimensions can either be found in the taxonomy of Enright and Gitomer (1989), or in the taxonomy of Oswald et al. (2004). The absence of the teamwork dimension in previous taxonomies might be explained by recent developments towards more group projects in the working environment, which simultaneously force universities to put more emphasis on it in their curricula.

Altogether the ETS PPI measures students' noncognitive academic potential formulated in a behavioral way. This is in line with the seminal paper of Wernimont and Campbell (1968) making the case for the measurement of past or present behavior in order to predict future behavior. Supervisor ratings of students' noncognitive academic potential might be especially valuable when predicting citizenship and counterproductive behavior, as they measure typical instead of maximal performance. Whereas standardized test results are a good predictor of maximal performance in terms of future grades, noncognitive predictors such as

Table 1
ETS® PPI scales and items

Knowledge and creativity
Has a broad perspective on the field
Is among the brightest persons I know
Produces novel ideas
Is intensely curious about the field
Communication skills
Speaks in a clear, organized, and logical manner
Writes with precision and style
Speaks in a way that is interesting
Organizes writing well
Teamwork
Supports the efforts of others
Behaves in an open and friendly manner
Works well in group settings
Gives criticism/feedback to others in a helpful way
Resilience
Accepts feedback without getting defensive
Works well under stress
Can overcome challenges and setbacks
Works extremely hard
Planning and organization
Sets realistic goals
Organizes work and time effectively
Meets deadlines
Makes plans and sticks to them
Ethics and integrity
Is among the most honest persons I know
Maintains high ethical standards
Is worthy of trust from others
Demonstrates sincerity

Note. From ETS® Personal Potential Index. ©2009 Educational Testing Service, Princeton, NJ. ETS® Personal Potential Index is a registered trademark of ETS.

the ETS PPI are expected to be more appropriate to predict typical performance operationalized as citizenship and counterproductive behavior. This is in line with the idea that every predictor (cognitive and noncognitive) has a “natural” equivalent on the criterion side (Marcus, Goffin, Johnston, & Rohstein, 2007). Thus, drawing on previous research claiming that noncognitive predictors are appropriate to predict citizenship and counterproductive behavior (McHenry, Hough, Toquam, Hanson, & Ashworth, 1990; Schmitt et al., 2003) we assume that:

H1. The ETS PPI is a valid predictor of students' academic potential in terms of (a) university citizenship behavior and (b) counterproductive academic behavior.

Furthermore, in order to justify the utility of supervisor ratings of students' academic potential in an applied context, incremental validity evidence above traditional predictors of study success is important (Hunsley & Meyer, 2003). Meta-analytic research shows that conscientiousness (Poropat, 2009), academic self-efficacy and achievement motivation (Robbins et al., 2004), grade goal (Richardson, Abraham, & Bond, 2012) and attitudinal constructs such as study habits (Credé & Kuncel, 2008) explain additional variance above and beyond traditional antecedents of study success. Oswald et al. (2004) also investigated the incremental validity of SJT and biodata measures. Both instruments provided incremental validity above and beyond traditional predictors. Notably however, both methods demonstrated stronger incremental validities when noncognitive aspects of performance such as citizenship and perseverance functioned as dependent variables. Meriac (2012) also provided evidence for the incremental validity of work ethic over traditional predictors of study success using university citizenship and academic counterproductive behavior as dependent variables. Therefore, we hypothesize the following:

H2. The ETS PPI provides incremental validity above and beyond U-GPA when university citizenship and counterproductive academic behavior function as dependent variables.

2. Method

2.1. Participants and procedure

A total of 278 graduate students enrolled at a Dutch University voluntarily participated in the project. The University is located close to the German and Belgium border. The institution has a diverse student body consisting of students from those three (the Netherlands, Germany, and Belgium) and other European and non-European countries. The educational teaching system makes the university well suited to collect ratings of citizenship and counterproductive behaviors. A key element of the problem based learning teaching system is that students learn in small, self-regulated groups of 10–14 persons. Proper group functioning is dependent on team processes, such as communication, knowledge sharing, and collaboration between students. As a consequence, contextual performance aspects like citizenship and deviant behavior is critical for the functioning of these groups.

Within a four-hour testing session, the participants (1) sent ETS PPI requests to up to five former supervisors, (2) completed the GRE® revised General Test, and (3) responded to a questionnaire that included demographic information (age, gender, faculty affiliation, and undergraduate grade point average). As an incentive for participation, students were entered into a random drawing and could win one of the ten iPads.

In order to send ETS PPI requests to former supervisors, students created an online ETS PPI account. They were instructed to indicate one to five former supervisors who they either knew from their bachelor studies (university context) or from previous internships or jobs (work context). Each supervisor received an automated email inviting them to access the web-based ETS PPI system to be able to complete the evaluation.

A total of 158 of the 278 participants created an account and asked, on average, 2.7 former supervisors to fill in the ETS PPI. Thus, in total 425 ETS PPI requests were sent out and a total of 176 completed ETS PPI evaluations were received (response rate of 43%). Overall 66% of the evaluators worked at universities and 34% were employed at companies or other institutions. The final sample, for which ETS PPI evaluations were available, consisted of 115 graduate students (female, $n = 61$, male, $n = 54$; age = $M: 23.85$, $SD: 3.00$) enrolled at five different faculties (business [$n = 34$], psychology [$n = 58$], law [$n = 11$], medicine [$n = 7$], arts and social science [$n = 5$]). Participating students had received their bachelor's degree in 23 different European (Netherlands, Ireland, Germany, United Kingdom, Czech Republic, Denmark, Belgium, Bulgaria, France, Spain) and Non-European countries (South Africa, Turkey, China, Ukraine, USA, Canada, Russia, Colombia, India, Indonesia, Iran, New Zealand, Vietnam).

Approximately three months after collection of predictor data, criterion data collection took place. To that end, participants were contacted and asked to complete a questionnaire assessing their university citizenship and counterproductive academic behavior tendencies. Data thus stems from two independent sources. ETS PPI evaluations were provided by former supervisors; citizenship and counterproductive behavior were assessed by the participants themselves.

2.2. Measures

2.2.1. The ETS PPI

The ETS PPI consists of six dimensions: knowledge and creativity, communication skills, teamwork, resilience, planning and organization, and ethics and integrity. Each dimension is assessed using four statements (see Table 1). A single item overall evaluation is also provided (“Given your evaluation of the applicant of the 6 personal dimensions, please provide an overall evaluation of the applicant.”). Evaluators rate the student on each statement comparing them to other students who have gone on to graduate study using a five-point scale: 1 = below

average, 2 = average, 3 = above average, 4 = outstanding (top 5%), and 5 = truly exceptional (top 1%). There is also an “insufficient opportunity to evaluate” field. Supervisors were instructed to rate participants in comparison to peer students within the same field of study.

As this study is one of the first validity studies on the ETS PPI overall, and the first validity study of the ETS PPI in Europe, we did a preliminary analysis to clarify the factor structure of the ETS PPI. As a first step, we inspected zero-order intercorrelations between ETS PPI dimensions revealing strong relationships ranging from $r = .59$ to $r = .84$ ($p < .01$; see Table 2). To clarify the factor structure, we tested a one-factor model and a second-order factor model with the help of the lavaan-package (Rosseel, 2012) included in the R environment. We conducted the analysis based on all ETS PPI evaluations that were available ($N = 176$). The second-order model was defined in such a way that a higher-order ETS PPI factor accounts for the pattern of relations between six correlated lower-order ETS PPI dimensions which respectively load on the four items that were designed to measure them.

Model fit was evaluated using the chi square (χ^2) test statistic, the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the standardized root mean square residual (SRMR). Results of the confirmatory factor analysis (CFA) revealed that the second-order factor model provided an acceptable model fit, χ^2 (248) = 448.28; RMSEA = .07, CFI = .94, SRMR = .05, whereas the one-factor model showed an unsatisfactory fit, χ^2 (275) = 1023.02; RMSEA = .12, CFI = .78, SRMR = .07. These findings imply that the six ETS PPI dimensions load on an overall factor. We accordingly operationalized the ETS PPI as an overall mean score for further analyses.

2.2.2. Undergraduate grade point average (U-GPA)

Participants provided their undergraduate grade point average. Because students received their bachelor’s degree in 23 different countries using different grading systems, they were also asked to indicate the highest and lowest possible grade underlying the grading system. Because the majority of the sample provided Dutch grades, we converted the remaining grades into the Dutch grading format (0 = lowest grade to 10 = highest grade). The grade 10.0 is hardly ever provided in the Dutch grading system (Nuffic, 2012). Therefore, the highest attainable grades in other national systems were converted into a 9.0. This conversion cannot truly equate grades assigned in different grading systems, but this rough approximation allows us to use data for students from all 23 countries in our sample.

2.2.3. University citizenship behavior

In order to measure university citizenship behavior, we chose a German scale (Gehring, 2006) that was developed based on research conducted in the work context (Organ, 1997; Staufenbiel & Hartz, 2000). We translated the scale into English using a translation/back-translation

process and selected 14 out of 27 items that were not specifically related to the German study context, but could be easily used in the Dutch university context as well. The scale measures four central dimensions of the citizenship behavior construct. Altruism is defined as helping behavior (e.g., “I help other students having problems with understanding the subject matter.”); Conscientiousness is characterized by being in time, or actively participating in group discussions (e.g., “I prepare for exams at an early stage.”); civic virtue is a voluntarily behavior to support activities at the university (e.g., “I constantly inform myself about developments in academic policy.”); and sportsmanship is characterized by accepting inconvenient study conditions without complaining about it (e.g., “When the content of a non-mandatory lecture or tutorial is important for me, I go there even when the time and the location is not suitable to me.”). Answers were provided on a five-point rating scale ranging from 1 = strongly disagree to 5 = strongly agree.

To examine the factor structure of the citizenship behavior construct we conducted a confirmatory factor analysis. We tested a one-factor model and a second-order model with one higher-order citizenship behavior factor accounting for the relation between four lower-order factors representing the differentiation in altruism (eight items), conscientiousness (two items), civic virtue (two items) and sportsmanship (two items). This is in line with current research (Hoffman, Blair, Meriac, & Woehr, 2007; Lepine, Erez, & Johnson, 2002) describing the citizenship behavior construct as a single factor model with underlying dimensions adding up to the latent construct. Whereas the fit of the one-factor model was unsatisfactory, χ^2 (90) = 175.02; RMSEA = .09, CFI = .68, SRMR = .10, the second-order factor model met the criteria for good model fit, χ^2 (76) = 93.11; RMSEA = .05, CFI = .94, SRMR = .07. With a value of .65, Cronbach’s alpha was somewhat low. However, Cronbach’s alpha has known limitations as a measure of internal consistency (Bentler, 2009; Revelle & Zinbarg, 2009). One well-known limitation is the fact that alpha is only a viable estimate of reliability when a measure is unidimensional (i.e., the correlation between the items is the same for all items). When the scale consists of some items that are more similar than others, alpha has been described as a poor estimate of the scale reliability. We therefore also estimated the reliability using an alternative measure of reliability – the greatest lower bound (glb; Woodhouse & Jackson, 1977) – that addresses this limitation of alpha. The glb has been recommended by Bentler (2009) and Sijtsma (2009) and is one of the several modern measures of reliability on the basis of one test administration. The glb for the university citizenship behavior scale was .83 indicating a satisfactory reliability.

2.2.4. Counterproductive academic behavior

In order to measure counterproductive academic behavior, we selected eight items from the established Workplace Deviance Scale (Bennett & Robinson, 2000) and adapted them to the academic context

Table 2
Descriptive statistics and intercorrelations between study variables.

	n	M	SD	Correlations											
				1	2	3	4	5	6	7	8	9	10		
1. KNOW	115	3.41	.71	–											
2. COM	115	3.35	.74	.84**	–										
3. TEAM	115	3.68	.72	.69**	.70**	–									
4. RES	115	3.55	.69	.74**	.71**	.68**	–								
5. PLAN	115	3.42	.80	.70**	.71**	.59**	.73**	–							
6. ETH	113	3.81	.75	.69**	.65**	.62**	.69**	.60**	–						
7. Overall	115	3.54	.64	.90**	.89**	.82**	.88**	.85**	.82**	–					
8. U-GPA	106	7.60	.71	.34**	.26**	.23**	.31**	.25*	.28**	.32**	–				
9. UCB	112	3.92	.42	.29**	.22*	.24*	.27**	.23*	.22*	.28**	.12	–			
10. CAB	112	3.13	.50	–.31**	–.24*	–.06	–.28**	–.36**	–.20*	–.29**	–.18	–.33**	–		

Note. KNOW = knowledge and creativity; COM = communication skills; TEAM = teamwork; RES = resilience; PLAN = planning and organization; ETH = ethics and integrity; Overall = ETS PPI mean score aggregated across evaluators; U-GPA = self-indicated undergraduate grade point average; UCB = university citizenship behavior; CAB = counterproductive academic behavior.

* $p < .05$.
** $p < .01$.

(e.g., “I take long breaks while studying.” or “I make offending comments about people studying or working at my university.”). Items that were too work specific (e.g., “I discussed confidential company information with an unauthorized person.” or “I falsified a receipt to get reimbursed for more money than I spent on business expenses.”) and did not have a meaningful equivalent in the academic context were dropped. The scale measures two facets of counterproductive behavior: Counterproductive behavior directed towards other individuals working or studying at the university (e.g., “I make fun of other students or people working at my university.”) and counterproductive behavior directed towards the university (e.g., “I am late or absent from study related meetings without permission or evident reason.”). All answers were provided on a five-point rating scale ranging from 1 = never to 5 = very often.

To examine the goodness of fit of the counterproductive academic behavior construct, we also conducted a confirmatory factor analysis in the R environment. We tested a one-factor model and a second-order factor model with a higher-order general factor and two correlated first-order factors representing counterproductive behaviors directed towards the university (CWB-O; four items) and counterproductive behaviors directed towards other individuals (CWB-I; four items). A second-order factor solution was supported, $\chi^2(22) = 28.61$; RMSEA = .05, CFI = .96, SRMR = .08, whereas a one-factor solution was rejected, $\chi^2(27) = 76.10$; RMSEA = .13, CFI = .68, SRMR = .11. These findings are in line with a recent meta-analysis (Berry, Ones, & Sackett, 2007) justifying the creation of an overall counterproductive behavior construct with CWB-I and CWB-O as underlying dimensions. The internal consistency of the eight item scale was .60 (Cronbach's alpha) and .80 (glb).

2.3. Data analysis

In order to investigate the predictive validity of the supervisor ratings of students' academic potential, we calculated zero-order correlations between all ETS PPI dimensions and the overall ETS PPI score with university citizenship and counterproductive academic behavior ratings. If more than one supervisor completed the ETS PPI, we calculated a mean score for every PPI item, and a mean score for every one of the

six PPI dimensions. The mean scores for the six dimensions were averaged in one overall ETS PPI score. To compute the incremental validity of the supervisor ratings above and beyond U-GPA, we conducted a series of hierarchical regression analyses, in order to investigate the predictive validity of the ETS PPI when using (a) the overall score, (b) each dimension separately, and (c) all dimensions together. In all of these analyses, U-GPA was entered as a control variable.

3. Results

Table 2 provides an overview of means, standard deviations, and intercorrelations between study variables, and shows that the overall ETS PPI score and all separate dimensions predict university citizenship behavior after a three month lag. The same holds true for counterproductive academic behavior except for the teamwork dimension.

Results of hierarchical regression analyses are presented in Table 3. In light of the high intercorrelations between ETS PPI dimensions and the results of the confirmatory factor analysis suggesting that the ETS PPI is best represented by one higher-order factor, our focus was on the overall ETS PPI score as predictor of university citizenship and counterproductive academic behavior. Results revealed that the overall score was a valid predictor of university citizenship and counterproductive academic behavior when the influence of U-GPA was controlled.

In addition to this main analysis, we conducted a series of supplementary analyses, including each ETS PPI dimension individually and including all ETS PPI dimensions jointly in a series of regression analyses. Yet, these findings may be interpreted with some caution because of the high intercorrelations between ETS PPI dimensions. Nevertheless, these supplementary analyses provide additional insights into the relationships of the individual ETS PPI subfacets with citizenship and counterproductive behavior. These analyses may be of some practical utility for universities focusing their selection procedure on specific aspects measured by the ETS PPI (e.g., communication skills) instead of simply choosing those students with the highest overall score. When entering ETS PPI dimensions individually in predicting citizenship behavior, the pattern of results was similar to the zero-order correlations depicted

Table 3
Hierarchical multiple regression analyses predicting university citizenship and counterproductive academic behavior from U-GPA and PPI results.

Predictor	ΔR^2	β														
<i>University citizenship behavior</i>																
Step 1	.02		.02		.02		.02		.02		.02		.02		.02	
U-GPA		.12		.12		.12		.12		.12		.12		.12		.12
Step 2	.08**		.11***		.05*		.06*		.06**		.06*		.04		.12*	
KNOW		–		.35***		–		–		–		–		–		.43*
COM		–		–		.23*		–		–		–		–		–.18
TEAM		–		–		–		.24*		–		–		–		.08
RES		–		–		–		–		.26**		–		–		.08
PLAN		–		–		–		–		–		.24*		–		.03
ETH		–		–		–		–		–		–		.19		–.09
Overall		.29**		–		–		–		–		–		–		–
Total R ²	.09**		.12***		.07*		.07*		.08**		.07*		.05		.14*	
<i>Counterproductive academic behavior</i>																
Step 1	.03		.03		.03		.03		.03		.03		.03		.03	
U-GPA		–.18		–.18		–.18		–.18		–.18		–.18		–.18		–.18
Step 2	.08**		.10***		.05*		.00		.08**		.13***		.04		.22***	
KNOW		–		–.33***		–		–		–		–		–		–.42*
COM		–		–		–.23*		–		–		–		–		.15
TEAM		–		–		–		–.05		–		–		–		.36**
RES		–		–		–		–		–.29**		–		–		–.11
PLAN		–		–		–		–		–		–.38***		–		–.36*
ETH		–		–		–		–		–		–		–.20		.05
Overall		–.30**		–		–		–		–		–		–		–
Total R ²	.11**		.13***		.09*		.04		.11**		.17***		.07		.26***	

Note. $N = 106$. U-GPA = self-indicated undergraduate grade point average; KNOW = knowledge and creativity; COM = communication skills; TEAM = teamwork; RES = resilience; PLAN = planning and organization; ETH = ethics and integrity; overall = ETS PPI mean score aggregated across evaluators.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

in Table 2. An exception is the ethics and integrity dimension which did not reach statistical significance when the influence of U-GPA was controlled. The same applied to counterproductive behavior. Next to the teamwork dimension that was not significantly related with counterproductive academic behavior, the ethics and integrity dimension was no longer statistically significant when controlling for U-GPA.

When all ETS PPI dimensions were included simultaneously in Step 2 of the regression analyses, knowledge and creativity was the only dimension that remained to be significant when predicting citizenship behavior. This is due to the high intercorrelations between individual dimensions. Similarly, only knowledge and creativity and planning and organization remained to be significantly negatively related to counterproductive behavior once all ETS PPI dimensions were included simultaneously. The significant positive estimate for teamwork is the result of a suppression effect as it only emerged when all ETS PPI dimensions were included simultaneously and not when they were entered individually.¹

4. Discussion

Results of the present study support the usefulness of supervisor ratings of students' academic potential as assessed in the ETS PPI to predict university citizenship and counterproductive academic behavior. This is in line with findings from the work context that supervisor' ratings are reliable predictors of job performance (Viswesvaran, Ones, & Schmidt, 1996). Furthermore, it broadens existing knowledge obtained in the academic context showing that evaluations provided by former supervisors are valid predictors of future grades (Aamodt, Bryan, & Whitcomb, 1993) and peer-ratings of academic performance (Stanton, Burstein, Kobos, & Loucks, 1979). U-GPA, however, did not explain a significant amount of variance within the outcome measures. Although meta-analyses have confirmed the predictive validity of U-GPA as indicator of study success (Kuncel, Credé, & Thomas, 2007; Kuncel, Hezlett, & Ones, 2001), it appears not to be a valid predictor of university citizenship and counterproductive academic behavior. Thus, results provide evidence for the incremental validity of the ETS PPI, as it predicts university citizenship and counterproductive academic behavior above and beyond U-GPA.

Supervisor ratings of students' academic potential as measured by using the ETS PPI will likely function as supplemental information rather than as a substitute for cognitive instruments. Thus, the incremental validity evidence we found in the present study is an important premise to justify their utility in an applied context. Combining multiple predictors has been described as a viable method to find a balance between validity and diversity of selection instruments (Sackett & Lievens, 2008). This idea is based on the assumption that noncognitive correlates, such as interests or personality, produce less subgroup differences compared to standardized tests that target cognitive skills (Hough, Oswald, & Ployhart, 2001; Sackett, Schmitt, Ellingson, & Kabin, 2001). Thus, in case admission committees aim to base their selection decisions on a broader conceptualization of academic performance in order to increase the diversity of the student body (Ployhart & Holtz, 2008; Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997), the use of standardized supervisor ratings seems to be an appropriate approach.

Previous studies investigating the incremental validity of noncognitive predictors primarily used students' grade point average as the dependent variable (Chapman & Hayslip, 2005; Farsides & Woodfield, 2003; Gore, 2006). Doing so, they found mixed results for the incremental validity of noncognitive instruments. The findings of the present study, however, underscore the need to extend the criterion space when investigating the incremental validity of noncognitive predictors to include aspects of academic performance that go beyond task performance. This is in line with Enright and Gitomer (1989) who argued that it is not sufficient to evaluate assessments in terms of their ability to predict future grades. Accordingly, we conclude that future validity studies

should not only extend the predictor, but also the criterion domain, in order to prevent underestimation of incremental validity evidence.

The validity evidence of the ETS PPI and the resulting possibility for its use in graduate selection has several practical implications. First of all, our study findings suggest that the ETS PPI is a valid and useful instrument for student selection. The ETS PPI may therefore be of great practical value to organizations as it is also efficient and time-effective for supervisors; supervisors only have to fill in one evaluation per student independent of the number of universities for which an applicant wants to apply. The student can then designate which universities the ETS PPI report is to be sent to and staff members will not have to write individual letters of recommendation or complete various forms required by the respective universities. Second, completing the ETS PPI is quicker than writing letters of recommendation, since the supervisor is responding to the same set of 24 statements across all students. Third, the standardization makes it easier for universities to compare students applying from different universities and even different countries. Students' mobility has considerably increased within Europe and applying students come from various countries. Thus, the standardized wording of the ETS PPI avoids possible misinterpretations due to cultural and language differences (Liu et al., 2009).

Taken together, findings of the present study are in line with Meriac (2012) who provided evidence for the positive relation between work ethic and university citizenship behavior, and the negative relation with counterproductive academic behavior. Although surprisingly the ethics and integrity dimension was not a valid predictor of counterproductive behavior when the influence of U-GPA was held stable, we provided evidence that overall supervisor ratings of students' academic potential as measured by the ETS PPI predict citizenship and counterproductive behavior in the academic context. Furthermore, we extended previous research (Meriac, 2012; Oswald et al., 2004; Zettler, 2011) by predicting student citizenship and counterproductive academic behavior using supervisor ratings instead of self-ratings.

4.1. Limitations and future directions

The present study is not without limitations; in the remainder, we critically discuss these limitations and identify potential areas of future research. First, the current study is based on a relatively small sample size. Future research might investigate the validity of the ETS PPI in a large-scale study. This would also allow comparisons to be made between different subsamples (e.g., female vs. male; European vs. Non-European) in order to evaluate the extent to which including the ETS PPI in a selection procedure reduces subgroup differences. Furthermore, the time between predictor and criterion data collection might be extended in future research, to investigate whether this has an influence on the validity of the ETS PPI.

Second, another shortcoming is that the present study was conducted in a low-stakes testing situation because students had already been admitted to their study program. This has both advantages and disadvantages: A disadvantage is that being aware that the student has already been accepted into a master's program may have caused biased supervisor ratings. It may therefore be valuable to replicate the present findings in a high-stakes testing situation. It would be especially interesting to compare findings from a high-stakes situation with the present study findings to investigate whether supervisors evaluate students systematically higher if students need the evaluation for an actual application. On the other hand, our procedure also had advantages as range restriction in the predictor is less of a concern. Our sample of students had not gone through a specific selection procedure before they could register for their master's program. Most faculties at the university used in the study required only verification of students' Bachelor degrees and a proof of their English proficiency level. As the ETS PPI was not part of this procedure we did not have range restriction in the predictor. In contrast, graduate programs in other countries are much more selective only admitting a small subset of applicants. Thus, predictive

¹ The same pattern of results emerged when the GRE as traditional predictor of study success was included in Step 1 of the regression analyses together with U-GPA.

validity studies conducted in more competitive educational systems might lead to lower levels of predictive validity due to range restriction underlying the need for further validity studies in other educational systems.

Third, one may consider the use of self-ratings of university citizenship and counterproductive academic behavior a shortcoming of the present study. However, research on citizenship and counterproductive behavior frequently relies on self-ratings (Fox, Spector, Goh, & Bruursema, 2007), because behaviors such as cheating or lying are mostly covert, and therefore remain unrecognized by others, making it difficult to assess counterproductive behavior with other-ratings. Recent meta-analyses comparing self- and other-ratings of citizenship and counterproductive behavior have also provided support for the utility of self-ratings (Berry, Carpenter, & Barratt, 2012; Carpenter, Berry, & Houston, 2014). Furthermore, the present study relied on multi-source data (former supervisor ratings for the ETS PPI and self-ratings for citizenship and counterproductive behavior) thereby reducing common method bias. In line with a study conducted in the work context (Connelly & Hülshager, 2012), future research might also consider the possibility to predict noncognitive aspects of performance by using other-ratings from outside the university context (e.g., family and friends).

Fourth and finally, a general shortcoming that is inherent in selection procedures such as letters of recommendations or the ETS PPI is the leniency bias (Aamodt, 1999). The likelihood that former supervisors provide favorable instead of honest ratings is high, as applicants themselves choose which former supervisors provide the information. In order to reduce this problem the answer scale of the ETS PPI was designed in such a way that it was imbalanced (Kyllonen, 2008). Assuming that supervisors provide their ratings on the positive side of the scale, this side was provided with a larger number of response choices: Three answer possibilities are above average and only one is below. Although empirical findings (Kyllonen, 2008) and the results of the present study show that supervisors are still prone to provide positive ratings (Kyllonen: $M = 3-4$; present study: $M = 3.57$), the ETS PPI allows for a more precise differentiation between highly-ranked applicants. However, future research is needed in order to further investigate ways to reduce the influence of leniency biases and increase fairness in student selection.

Acknowledgments

This research was supported by a grant from Educational Testing Service (ETS), Princeton, New Jersey. ETS provided access to the ETS® PPI and the GRE® revised General Test. ETS and GRE are registered trademarks of Educational Testing Service (ETS) in the United States and other countries.

References

- Aamodt, M. G. (1999). *Applied industrial organizational psychology* (3rd ed.). London: Brooks/Cole.
- Aamodt, M. G., Bryan, D. A., & Whitcomb, A. J. (1993). Predicting performance with letters of recommendation. *Public Personnel Management*, 22, 81–90 (Retrieved from: [http://search.ebscohost.com/login.aspx?direct=true&db=buh&bquery=\(SO+\(Public+personnel+management\)\)+AND+\(DT+1993\)+AND+\(TI+\(predicting+performance+with+letters+%26quot%3b%26quot%3b+recommendation\)\)&type=1&site=ehost-live&scope=site](http://search.ebscohost.com/login.aspx?direct=true&db=buh&bquery=(SO+(Public+personnel+management))+AND+(DT+1993)+AND+(TI+(predicting+performance+with+letters+%26quot%3b%26quot%3b+recommendation))&type=1&site=ehost-live&scope=site)).
- Bennett, R. J., & Robinson, S. L. (2000). Development of a measure of workplace deviance. *Journal of Applied Psychology*, 85, 349–360. <http://dx.doi.org/10.1037/0021-9010.85.3.349>.
- Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, 74, 137–143. <http://dx.doi.org/10.1007/S11336-008-9100-1>.
- Berry, C. M., Carpenter, N. C., & Barratt, C. L. (2012). Do other-reports of counterproductive work behavior provide an incremental contribution over self-reports? A meta-analytic comparison. *Journal of Applied Psychology*, 97, 613–636. <http://dx.doi.org/10.1037/a0026739>.
- Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology*, 92, 410–424. <http://dx.doi.org/10.1037/0021-9010.92.2.410>.
- Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance*, 10, 99–109. http://dx.doi.org/10.1207/s15327043hup1002_3.
- Briel, J., Bejar, I., Chandler, M., Powell, G., Manning, K., Robinson, D., et al. (2000). *GRE horizons planning initiative. (unpublished graduate record examination report)*. Princeton, NJ: Educational Testing Service.
- Carpenter, N. C., Berry, C. M., & Houston, L. (2014). A meta-analytic comparison of self-reported and other-reported organizational citizenship behavior. *Journal of Organizational Behavior*, 35, 547–574. <http://dx.doi.org/10.1002/job.1909>.
- Chapman, B. P., & Hayslip, B., Jr. (2005). Incremental validity of a measure of emotional intelligence. *Journal of Personality Assessment*, 85, 154–169. http://dx.doi.org/10.1207/s15327752jpa8502_08.
- Connelly, B.S., & Hülshager, U. R. (2012). A narrower scope or a clearer lens for personality? Examining sources of observers' advantages over self-reports for predicting performance. *Journal of Personality*, 80, 603–631. <http://dx.doi.org/10.1111/j.1467-6494.2011.00744.x>.
- Connelly, B.S., & Ones, D. S. (2010). Another perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, 136, 1092–1122. <http://dx.doi.org/10.1037/a0021212>.
- Credé, M., & Kuncel, N. R. (2008). Study habits, skills, and attitudes: The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science*, 3, 425–453. <http://dx.doi.org/10.1111/j.1745-6924.2008.00089.x>.
- Credé, M., & Niehorster, S. (2009). Individual difference influences on self-focused and other-focused counterproductive student behaviors. *Personality and Individual Differences*, 47, 769–776. <http://dx.doi.org/10.1016/j.paid.2009.06.018>.
- Enright, M. K., & Gitomer, D. (1989). *Toward a description of successful graduate students. (GREB 85-17R; RR 89-09)*. Princeton, NJ: Educational Testing Service (Retrieved from <http://origin-www.ets.org/Media/Research/pdf/RR-89-09-Enright.pdf>).
- Farsides, T., & Woodfield, R. (2003). Individual differences and undergraduate academic success: The roles of personality, intelligence, and application. *Personality and Individual Differences*, 34, 1225–1243. [http://dx.doi.org/10.1016/S0191-8869\(02\)00111-3](http://dx.doi.org/10.1016/S0191-8869(02)00111-3).
- Fox, S., Spector, P. E., Goh, A., & Bruursema, K. (2007). Does your coworker know what you're doing? Convergence of self- and peer-reports of counterproductive work behavior. *International Journal of Stress Management*, 14, 41–60. <http://dx.doi.org/10.1037/1072-5245.14.1.41>.
- Gehring, F. (2006). University citizenship behavior. Development and validation of an inventory on contextual performance at universities (Unpublished diploma thesis). Aachen, Germany: RWTH Aachen University.
- Gore, P. A. (2006). Academic self-efficacy as a predictor of college outcomes: Two incremental validity studies. *Journal of Career Assessment*, 14, 92–115. <http://dx.doi.org/10.1177/1069072705281367>.
- Hoffman, B. J., Blair, C. A., Meriac, J. P., & Woehr, D. J. (2007). Expanding the criterion domain? A quantitative review of the OCB literature. *Journal of Applied Psychology*, 92, 555–566. <http://dx.doi.org/10.1037/0021-9010.92.2.555>.
- Hough, L. M., Oswald, F., & Ployhart, R. (2001). Determinants, detection, and amelioration of adverse impact in personnel selection procedures: Issues, evidence, and lessons learned. *International Journal of Selection and Assessment*, 9, 152–194. <http://dx.doi.org/10.1111/1468-2389.00171>.
- Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment*, 15, 446–455. <http://dx.doi.org/10.1037/1040-3590.15.4.446>.
- Kim, S., & Kyllonen, P. C. (2007). Rasch measurement in developing faculty ratings of students applying to graduate school. *Journal of Applied Measurement*, 9, 168–181 (Retrieved from <http://europepmc.org/abstract/MED/18480513>).
- Kuncel, N. R., Credé, M., & Thomas, L. L. (2007). A comprehensive meta-analysis of the predictive validity of the Graduate Management Admission Test (GMAT) and undergraduate grade point average (UGPA). *Academy of Management Learning and Education*, 6, 51–68. <http://dx.doi.org/10.5465/AMLE.2007.24401702>.
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, 127, 162–181. <http://dx.doi.org/10.1037/0033-2909.127.1.162>.
- Kyllonen, P. C. (2008). *The research behind the ETS Personal Potential Index (PPI)*. NY, ETS: Princeton (Retrieved from http://www.ets.org/Media/Products/PPI/10411_PPI_bkgrd_report_RD4.pdf).
- Kyllonen, P. C., & Kim, S. (2005, April). Personal qualities in higher education: Dimensionality of faculty ratings of students applying to graduate school. *Paper presented at the annual meeting of the American Educational Research Association, Montréal, Canada*.
- Lepine, J. A., Erez, A., & Johnson, D. E. (2002). The nature and dimensionality of organizational citizenship behavior: A critical review and meta-analysis. *Journal of Applied Psychology*, 87, 52–65. <http://dx.doi.org/10.1037/0021-9010.87.1.52>.
- Lievens, F., Buysse, T., & Sackett, P. R. (2005). The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology*, 90, 442–452. <http://dx.doi.org/10.1037/0021-9010.90.3.442>.
- Lievens, F., & Coetsier, P. (2002). Situational tests in student selection: An examination of predictive validity, adverse impact, and construct validity. *International Journal of Selection and Assessment*, 10, 245–257. <http://dx.doi.org/10.1111/1468-2389.00215>.
- Liu, O. L., Minsky, J., Ling, G., & Kyllonen, P. (2009). Using the standardized letters of recommendation in selection: Results from a multidimensional Rasch model. *Educational and Psychological Measurement*, 69, 475–492. <http://dx.doi.org/10.1177/0013164408322031>.
- Marcus, B., Goffin, R. D., Johnston, N. G., & Rothstein, M. G. (2007). Personality and cognitive ability as predictors of typical and maximum managerial performance. *Human Performance*, 20, 275–285. <http://dx.doi.org/10.1080/08959280701333362>.

- McCabe, D. L., Butterfield, K. D., & Trevino, L. K. (2006). Academic dishonesty in graduate business programs: Prevalence, causes, and proposed action. *Academy of Management Learning and Education*, 5, 294–305. <http://dx.doi.org/10.5465/AMLE.2006.22697018>.
- McCarthy, J. M., & Goffin, R. D. (2001). Improving the validity of letters of recommendation: An investigation of three standardized reference forms. *Military Psychology*, 13, 199–222. http://dx.doi.org/10.1207/S15327876MP1304_2.
- McDaniel, M.A., Morgeson, F. P., Finnegan, E. B., Campion, M.A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, 86, 730–740. <http://dx.doi.org/10.1037/0021-9010.86.4.730>.
- McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M.A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology*, 43, 335–354. <http://dx.doi.org/10.1111/j.1744-6570.1990.tb01562.x>.
- Meriac, J. P. (2012). Work ethic and academic performance: Predicting citizenship and counterproductive behavior. *Learning and Individual Differences*, 22, 549–553. <http://dx.doi.org/10.1016/j.lindif.2012.03.015>.
- Morphew, C. C., & Hartley, M. (2006). Mission statements: A thematic analysis of rhetoric across institutional type. *The Journal of Higher Education*, 77, 456–471. <http://dx.doi.org/10.1353/jhe.2006.0025>.
- Ng, T. W., & Feldman, D. C. (2009). How broadly does education contribute to job performance? *Personnel Psychology*, 62, 89–134. <http://dx.doi.org/10.1111/j.1744-6570.2008.01130.x>.
- Nicklin, J. M., & Roch, S. G. (2009). Letters of recommendation: Controversy and consensus from expert perspectives. *International Journal of Selection and Assessment*, 17, 76–91. <http://dx.doi.org/10.1111/j.1468-2389.2009.00453.x>.
- Nuffic (2012). *Country module – the Netherlands. Degrees and qualifications in the Netherlands*, 3. (pp. 1–39) (Retrieved from <https://www.nuffic.nl/en/library/country-module-netherlands.pdf>).
- Organ, D. W. (1997). Organizational citizenship behavior: It's construct clean-up time. *Human Performance*, 10, 85–97. http://dx.doi.org/10.1207/s15327043hup1002_2.
- Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M.A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology*, 89, 187–207. <http://dx.doi.org/10.1037/0021-9010.89.2.187>.
- Pérez-Peña, R. (2013, February 1). Students disciplined in Harvard scandal. *The New York Times* (Retrieved from http://www.nytimes.com/2013/02/02/education/harvard-forced-dozens-to-leave-in-cheating-scandal.html?_r=0).
- Ployhart, R. E., & Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing race/ethnicity and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, 61, 153–172. <http://dx.doi.org/10.1111/j.1744-6570.2008.00109.x>.
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, 135, 322–338. <http://dx.doi.org/10.1037/a0014996>.
- Poropat, A. E. (2011). The role of citizenship performance in academic achievement and graduate employability. *Education and Training*, 53, 499–514. <http://dx.doi.org/10.1108/00400911111159467>.
- Ramsay, L., Schmitt, N., Oswald, F. L., Kim, B. H., & Gillespie, M. (2006). The impact of situational context variables on responses to biodata and situational judgment inventory items. *Psychology Science*, 48, 268–287 (Retrieved from <http://io.psy.msu.edu/cbstudy/Ramsay%20SJI%20-%20Psychology%20Sci%202006.pdf>).
- Range, L. M., Menyhart, A., Walsh, M. L., Hardin, K. N., Ellis, J. B., & Craddock, R. (1991). Letters of recommendation: Perspectives, recommendations, and ethics. *Professional Psychology: Research and Practice*, 22, 389–392. <http://dx.doi.org/10.1037/0735-7028.22.5.389>.
- Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, 74, 145–154. <http://dx.doi.org/10.1007/s11336-008-9102-z>.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138, 353–387. <http://dx.doi.org/10.1037/a0026838>.
- Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130, 261–288. <http://dx.doi.org/10.1037/0033-2909.130.2.261>.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36 (Retrieved from: http://www.lce.esalq.usp.br/arquivos/aulas/2013/encontro_ppg/Lucia/paper.pdf).
- Rotundo, M., & Sackett, P. R. (2002). The relative importance of task, citizenship, and counterproductive performance for supervisor ratings of overall performance: A policy capturing study. *Journal of Applied Psychology*, 87, 66–80. <http://dx.doi.org/10.1177/014920639402000307>.
- Sackett, P. R., & Lievens, F. (2008). Personnel selection. *Annual Review of Psychology*, 59, 419–450. <http://dx.doi.org/10.1146/annurev.psych.59.103006.093716>.
- Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist*, 56, 302–318. <http://dx.doi.org/10.1037/0003-066X.56.4.302>.
- Schmitt, N. (2012). Development of rationale and measures of noncognitive college student potential. *Educational Psychologist*, 47, 18–29. <http://dx.doi.org/10.1080/00461520.2011.610680>.
- Schmitt, N., Keeney, J., Oswald, F. L., Pleskac, T., Quinn, A., Sinha, R., et al. (2009). Prediction of 4-year college student performance using cognitive and noncognitive predictors and the impact of demographic status on admitted students. *Journal of Applied Psychology*, 94, 1479–1497. <http://dx.doi.org/10.1037/a0016810>.
- Schmitt, N., Oswald, F. L., Friede, A., Imus, A., & Merritt, S. (2008). Perceived fit with an academic environment: Attitudinal and behavioral outcomes. *Journal of Vocational Behavior*, 72, 317–335. <http://dx.doi.org/10.1016/j.jvb.2007.10.007>.
- Schmitt, N., Oswald, F. L., Kim, B. H., Gillespie, M.A., Ramsay, L. J., & Yoo, T. (2003). Impact of elaboration on social desirability and the validity of biodata measures. *Journal of Applied Psychology*, 88, 979–988. <http://dx.doi.org/10.1037/0021-9010.88.6.979>.
- Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology*, 82, 719–730. <http://dx.doi.org/10.1037/0021-9010.82.5.719>.
- Sijtsma, K. (2009). Reliability beyond theory and into practice. *Psychometrika*, 74, 169–173. <http://dx.doi.org/10.1007/s11336-008-9103-y>.
- Sinha, R., Oswald, F., Imus, A., & Schmitt, N. (2011). Criterion-focused approach to reducing adverse impact in college admissions. *Applied Measurement in Education*, 24, 137–161. <http://dx.doi.org/10.1080/08957347.2011.554605>.
- Stanton, B. C., Burstein, A. G., Kobos, J. C., & Loucks, S. (1979). The dean's letter of recommendation and resident performance. *Academic Medicine*, 54, 812–813. <http://dx.doi.org/10.1097/00001888-197910000-00009>.
- Staufenbiel, T., & Hartz, C. (2000). Organizational citizenship behavior: Development and validation of a measurement instrument. *Diagnostica*, 46, 73–83. <http://dx.doi.org/10.1026/0012-1924.46.2.73>.
- Verduin, J. R., & Clark, T. A. (1991). *Distance education: The foundations of effective practice*. San Francisco, CA: Jossey-Bass Publishers.
- Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment*, 8, 216–226. <http://dx.doi.org/10.1111/1468-2389.00151>.
- Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, 81, 557–574. <http://dx.doi.org/10.1037/0021-9010.81.5.557>.
- Walpole, M. B., Burton, N. W., Kanyi, K., & Jackenthal, A. (2001). *Selecting successful graduate students in-depth interviews with GRE users*. Princeton, NJ: Educational Testing Service (Retrieved from <http://www.ets.org/Media/Research/pdf/RR-02-08-Walpole.pdf>).
- Walters, A.M., Kyllonen, P. C., & Plante, J. W. (2006). Developing a standardized letter of recommendation. *Journal of College Admission*, 191, 8–17 (Retrieved from <http://files.eric.ed.gov/fulltext/EJ741520.pdf>).
- Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology*, 52, 372–376. <http://dx.doi.org/10.1037/h0026244>.
- Woodhouse, B., & Jackson, P. H. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: II: A search procedure to locate the greatest lower bound. *Psychometrika*, 42, 579–591. <http://dx.doi.org/10.1007/BF02295980>.
- Zettler, I. (2011). Self-control and academic performance: Two field studies on university citizenship behavior and counterproductive academic behavior. *Learning and Individual Differences*, 21, 119–123. <http://dx.doi.org/10.1016/j.lindif.2010.11.002>.