Graduate Student Selection: Graduate record examination, socioeconomic status, and undergraduate grade point average as predictors of study success in a western European University

Inge T. L. Schwager*, Ute R. Hülsheger**, Brent Bridgeman*** and Jonas W. B. Lang****

*Department of Work and Social Psychology, Maastricht University, Universiteitssingel 40, 6229 ES Maastricht, The Netherlands. inge.schwager@maastrichtuniversity.nl
**Department of Work and Social Psychology, Maastricht University, Maastricht, The Netherlands
***Educational Testing Service, Princeton, NJ, USA
****Department of Personnel Management, Work and Organizational Psychology, Ghent University, Ghent, Belgium

Graduate students' mobility has increased within Europe. Yet, empirical evidence on the validity of standardized admission tests in Europe is still scarce. The aim of the present study was therefore to investigate the incremental validity of the *GRE*® revised general test above undergraduate grade point average (U-GPA) by focusing on a multinational sample of master students (*N* = 282) enrolled at a Dutch university. Results indicated that the Analytical Writing part of the GRE predicts graduate grade point average above and beyond U-GPA. Furthermore, the results suggest that the relationship is independent of students' socioeconomic status as indexed by parental education. Implications for graduate student selection in a European context are discussed.

1. Introduction

European universities increasingly try to attract international students especially on the graduate level. In doing so, they intend to improve the quality of higher education, to encourage knowledge exchange, and to stimulate the national labor market (Becker & Kolster, 2012). Furthermore, national policy makers seek to strengthen the so-called pull factors, which determine universities' attractiveness for foreign students (Mazzarol & Soutar, 2002). More and more graduate students take the opportunity to study abroad in order to gain international experience and to boost their employability. As a consequence, student mobility has increased considerably over the recent years in Europe.

With increasing student mobility, universities face the necessity to use cost-effective, fair, and valid methods of student selection in order to evaluate such a diverse body of applicants and to enhance the likelihood of admitting students that are well suited and truly have the potential to contribute to the aforementioned objectives.

Standardized tests such as the Graduate Record Examination (GRE) are a reasonable alternative to the frequent use of undergraduate grade point average (U-GPA) in selecting graduate students (Kuncel & Hezlett, 2007). In contrast to U-GPA, test scores resulting from the three GRE parts measuring Verbal, Quantitative, and Analytical abilities can be easily compared among large numbers of applicants with different educational and national backgrounds. Recent meta-analyses have provided evidence for the predictive validity of such standardized admission tests (Kuncel, Credé, & Thomas, 2007;

Kuncel, Hezlett, & Ones, 2001; Kuncel, Wee, Serafin, & Hezlett, 2010; Powers, 2004). Authors of the largest meta-analysis including 1,521 studies (Kuncel et al., 2001) provided evidence for the validity of the GRE general test in predicting G-GPA, faculty ratings of performance, and comprehensive examination scores.

However, the meta-analyses cited above (Goldberg & Alliger, 1992; Kuncel et al., 2001; Kuncel et al., 2010; Morrison & Morrison, 1995) are based predominantly on criterion data from the United States. Criterion data for the European context are limited. We are only aware of research that follows up the criterion validity of tests used for undergraduate selection up to the graduate level (Hell, Trapmann, & Schuler, 2007; Lievens, Buyse, & Sackett, 2005). However, studies providing criterion data for standardized tests specifically aimed at selecting graduate students into European institutions after they complete their undergraduate degree are scarce. Yet, mobility of graduate students in Europe is greatly increasing in recent years (e.g. Statistics Netherlands, 2010), and consequently, graduate student selection is a major policy problem in many European countries.

Validity evidence obtained in the US context may not be generalizable to the European context without further testing, as cultural differences and differences in educational systems may affect validities of selection methods (Hülsheger, Maier, & Stumpp, 2007; Salgado & Anderson, 2002). For instance, the predictive validity of the GRE may be affected by different grading traditions: European education systems differ from the US system with respect to grading scheme and grading severity. The Netherlands, for example, has a grading culture that is much stricter than the US grading culture as the highest possible grade (a 10 on a 1–10 scale) is hardly ever awarded in the Netherlands (Nuffic, 2006). Furthermore, educational systems are characterized by their degree of differentiation and tracking at the secondary school level. In stratified school systems, students are sorted into different types of schools preparing them for different educational and occupational trajectories according to their level of achievement. In less stratified educational systems, comprehensive secondary schools are common and ability-based tracking is less widespread (Kerckhoff, 2001). Stratification has an influence on the homogeneity of the student body in different countries, which, in turn, might influence the predictive validity of standardized tests.

The first goal of the present study is to investigate the validity of the GRE to predict study success at the graduate level. This paper reports the first GRE validation study conducted in a European country that we are aware of. To measure study success, we consider not only graduate grade point average (G-GPA) but also the degree attainment within scheduled time frame. Whereas G-GPA is one of the most widely used measures of graduate study success (see Kuncel et al., 2001), degree attainment within the scheduled time frame defined in the curriculum is of special interest because universities need to estimate the resources that need to be allocated on every single student based on predefined study duration. In case many students exceed this time span, additional resources are needed, which, in turn, can have a negative impact on universities' capacities. In analyzing GRE validities for these outcomes, we control for U-GPA in order to assure that the GRE is a valid predictor of study success above and beyond U-GPA. U-GPA is widely used as a selection instrument in Europe. From a utility perspective, it is therefore important to investigate the incremental validity of the GRE above this well-established predictor to make sure that an added value is obtained when using the GRE for student selection.

A second goal of the present paper is to investigate whether the validity of the GRE in predicting criteria of study success at a European university is not an artifact of socioeconomic status (SES). Critics of post-secondary admissions tests like the GRE have asserted that relationships between standardized admissions tests and criteria of study success are only an artifact of parental SES in terms of parents' earnings or education (e.g. Crosby, Iyer, Clayton, & Downing, 2003; Kohn, 2001; Zwick, 2004). This assertion rests on the idea that SES influences both grades and test results in admission tests, such that any association found between admission tests and grades could be explained by SES and therefore not be indicative of the predictive validity of the test itself. Notably, this criticism also includes the idea that a positive association between SES and test scores does not reflect a true higher standing of high SES students on what admission tests measure but simply higher familiarity with test-taking. Similarly, the criticism implies that any association found between SES and study grades is not caused by higher true academic performance for students high in SES but asserts that it is an artifact of grading being biased in favor of high SES students (for a detailed discussion, see Sackett, Borneman, & Connelly, 2008; Sackett, Kuncel, Arneson, Cooper, & Waters, 2009). Following this criticism, any relation found between test scores and criteria of study success would be an artifact of SES and should therefore disappear when one controls for parental SES. Sackett et al. (2009) tested this idea in a comprehensive study including various large-scale US datasets on the SAT (formerly Scholastic Aptitude Test). They found that although SES was related to SAT scores, the SAT–grade relationship was only marginally reduced when the influence of SES was controlled for. The present study sets out to replicate Sackett et al.'s (2009) findings in a European sample and using a graduate admission test (GRE) rather than an undergraduate admission test (SAT).

2. Method

2.1. Participants and procedure

Two hundred eighty-two international students (female = 155; male = 127; age: $M = 23.55$, $SD = 2.36$) enrolled at five faculties (Business [$n = 83$], Psychology [$n = 125$], Law [$n = 49$], Medicine [$n = 16$], and Arts and Social Science [$n = 9$]) at a Dutch University volunteered for the project at the beginning of their master program. They indicated that they originated from 22 European (e.g., the Netherlands, Germany, Bulgaria, Poland, Lithuania, Belgium, Spain), and 15 non-European countries (e.g., China, Turkey, Ukraine, Canada, Mexico, Colombia, Vietnam). Furthermore, they received their bachelor degrees from 35 different countries and were enrolled in English-speaking master programs. This is in line with the assumption that Dutch universities have a strong international focus due to the central location within Europe. Furthermore, numerous master's degree programs are taught in English in order to reduce potential language barriers (Becker and Kolster, 2012) and to strengthen the attractiveness of Dutch universities for foreign students (Mazzarol & Soutar, 2002). The majority of participating students were enrolled in 1-year master programs; only six participants were enrolled in a 2-year research master program.

The study had a longitudinal set-up, with GRE testing taking place at the beginning of participants' master studies and criterion data collection taking place approximately two years later. At time 1, master students participated in a 4-hr testing session, in which they completed the revised (now current) version of the GRE, and a questionnaire including demographic information (age, gender, faculty affiliation, SES, U-GPA).

Criterion data were collected at time 2, two years after completion of the GRE. Specifically, information on students' G-GPA and degree attainment was obtained from student registry offices. At time 2, G-GPAs were available for 236 participants (84%). The remaining participants had not finished their master program until this date. For our second criterion, degree attainment within scheduled time frame, the entire sample ($N = 282$) has been taken into account.

2.2. Measures

2.2.1. Graduate record examination (GRE) revised general test

We used the revised GRE general test, consisting of three parts: GRE-V (Verbal Reasoning), GRE-Q (Quantitative Reasoning), and GRE-AW (Analytical Writing). Whereas GRE-V measures the ability to analyze and evaluate written material and GRE-Q measures basic mathematical skills and the understanding of mathematical concepts, GRE-AW measures critical thinking and analytical writing skills. GRE-V and GRE-Q scores are reported on a scale from 130 and 170 in 1-point increments. For GRE-AW, the score range is from 0 to 6 in half-point increments.

2.2.2. Socioeconomic status (SES)

Participants were asked to indicate the highest educational degree of their parents, respectively. They had to choose one of six options: 1 = *no school degree*; 2 = *basic school degree*; 3 = *advanced school degree*; 4 = *vocational degree*; 5 = *basic university degree*; 6 = *advanced university degree*; 7 = *doctoral degree*. A mean score for mothers' and fathers' education was calculated.

2.2.3. Undergraduate grade point average (U-GPA)

We asked the participants to indicate their U-GPA. Since the majority of the sample provided Dutch grades, we converted the remaining grades into the Dutch grading format (from 0 = lowest grade to 10 = highest grade). Given that 10s are hardly ever provided in the Dutch grading system (Nuffic, 2006; 6 = 33–37%, 7 = 33–37%, 8 = 22–26%, 9 = 6%, 10 = 1%), the highest attainable grades in other national systems were converted into a 9.0.

2.2.4. G-GPA and degree attainment within scheduled time frame

Faculties' student registry offices provided us with participant's G-GPA and start and graduation date. In case students attained their degree within the scheduled time frame (in most cases, 1 year), the outcome variable was coded with 1. In case it took students longer to finish their master program, the variable was coded with 0.

2.3. Data analysis

To investigate GRE's predictive validity, we calculated zero-order correlations between GRE-V, GRE-Q, and GRE-AW and the two outcome measures (G-GPA and degree attainment within scheduled time frame). Grades provided over the course of several months by different supervisors affiliated with different faculties are prone to rating errors. One may therefore wish to correct for the fact that measurement error in the criterion may lead to lower observed correlations between GRE and G-GPA, thereby leading to an underestimation of the true validity of the GRE. In accordance with previous studies investigating the predictive validity of standardized admission tests (Kuncel et al., 2007; Kuncel et al., 2001; Kuncel et al., 2010), we therefore corrected the observed GRE-G-GPA correlations for attenuation due to measurement error in the criterion (Hunter & Schmidt, 2004). In order to get an estimation of the reliability of G-GPA, we computed the reliability coeffi-

cient between the mean grade students received for their course work and the grade they received for their master thesis project. The course work grade is the average grade students obtain for four different courses providing general knowledge in their corresponding master specialization. In order to finish their studies, students have to conduct an empirical research project based on which they write their master thesis. The course work and the master thesis project are equally important with respect to earned credits and sum up into the final average grade. The raw correlation of .38 ($p < .01$) between the practical part and the research-based part resulted in a reliability of .54 for the G-GPA estimate. Accordingly, we divided the observed correlations between the independent variables and G-GPA by the square root of the reliability coefficient (Hunter & Schmidt, 2004) in order to obtain an estimate of the operational validity of the GRE. For degree attainment within scheduled time frame, no correction was made as this is an objective criterion that is not affected by measurement error.

In order to examine the influence of parental SES and the incremental validity of the GRE above and beyond U-GPA when predicting G-GPA, we used hierarchical regression analysis. For G-GPA as dependent variable, we ran two regression analyses – one based on the raw data and one based on the correlation matrix including the correlations corrected for criterion unreliability. Since degree attainment within scheduled time frame was a dichotomous variable, we used a logistic regression analysis to analyze the data. In step 1, we controlled for SES; in step 2, we entered U-GPA; and in step 3, we included GRE-V, GRE-Q, and GRE-AW in both regression analyses.

3. Results

Table 1 provides an overview about the zero-order correlations between the predictor-criterion combinations in order to provide information about the predictive validity of the GRE. Considering the entire sample, GRE-V, GRE-Q, and GRE-AW were all significantly related to subsequent G-GPA. In contrast, none of the three GRE subscales were significantly related to whether students attained their master's degree within the scheduled time frame.

Table 2 provides the results for the incremental validity of the GRE above and beyond U-GPA when G-GPA functioned as dependent variables. Considering results based on uncorrected correlations, findings reveal that although U-GPA was a valid predictor of G-GPA

Table 1. Intercorrelations between GRE results, U-GPA, SES and outcome variables

	n	M	SD	Correlations						
				1	2	3	4	5	6	7
				Complete sample						
1. GRE-V	278	146.80	6.91	–						
2. GRE-Q	278	147.71	7.56	.40**	–					
3. GRE-AW	279	3.11	0.63	.42**	.30**	–				
4. SES	279	4.76	1.18	.05	.05	.02	–			
5. U-GPA	257	7.61	0.67	.13*	.10	.20**	.16*	–		
6. G-GPA	236	7.53	0.53	.21** (.28)	.17** (.23)	.31** (.42)	.19** (.26)	.22** (.30)	–	
7. Degree attainment	282	0.56	0.50	.03	.09	−.01	.03	.14*	.20**	–
				Psychology student sample						
1. GRE-V	124	145.90	6.47	–						
2. GRE-Q	124	144.83	6.02	.43**	–					
3. GRE-AW	125	3.03	0.71	.46**	.30**	–				
4. SES	125	4.81	1.22	.03	.05	.09	–			
5. U-GPA	115	7.55	0.63	.27**	.12	.37**	.08	–		
6. G-GPA	97	7.66	0.50	.25* (.40)	.24* (.38)	.42** (.67)	.14 (.22)	.45** (.72)	–	
7. Degree attainment	126	0.62	0.49	.05	.14	.02	−.06	.21*	.02	–
				Business student sample						
1. GRE-V	83	147.83	6.88	–						
2. GRE-Q	83	153.19	7.46	.42**	–					
3. GRE-AW	83	3.18	0.52	.34**	.23*	–				
4. SES	83	4.75	1.14	.13	.26*	−.14	–			
5. U-GPA	77	7.59	0.66	.24*	.51**	.17	.26*	–		
6. G-GPA	79	7.48	0.54	.29* (.37)	.24* (.30)	.28* (.35)	.25* (.32)	.30* (.38)	–	
7. Degree attainment	83	0.52	0.50	.13	.26*	.08	.07	.27*	.26*	–

Notes: * $p < .05$; ** $p < .01$. GRE-V = GRE Verbal Reasoning; GRE-Q = GRE Quantitative Reasoning; GRE-AW = GRE Analytical Writing; SES = mean score of father's and mother's education; U-GPA = undergraduate grade point average; G-GPA = graduate grade point average. Degree attainment within scheduled time frame was coded 1 = finished in time; 0 = finished later than scheduled time frame. Correlations corrected for criterion unreliability are presented in parentheses.

Table 2. Hierarchical multiple regression analyses predicting G-GPA from SES, UGPA, and GRE results

Predictor	G-GPA	
	ΔR^2	β
Step 1	.02* (.07)	
SES		.15* (.26)
Step 2	.03* (.05)	
U-GPA		.18** (.20)
Step 3	.09*** (.15)	
GRE-V		.09 (.08)
GRE-Q		.03 (.07)
GRE-AW		.24** (.36)
Total R^2	.14*** (.27)	

Notes: * $p < .05$; ** $p < .01$; *** $p < .001$. $n = 236$. G-GPA = graduate grade point average; SES = mean score of father's and mother's education; U-GPA = undergraduate grade point average; GRE-V = GRE Verbal Reasoning; GRE-Q = GRE Quantitative Reasoning; GRE-AW = GRE Analytical Writing. Coefficients corrected for criterion unreliability are presented in parentheses.

Table 3. Logistic regression analysis predicting degree attainment within scheduled time frame from SES, U-GPA, and GRE

Predictor	Model 1	Model 2	Model 3
SES	.01	−.03	−.04
U-GPA		.43*	.45*
GRE-V			.01
GRE-Q			.02
GRE-AW			−.23
Model fit			
χ^2	0.02	4.82	6.75
df	1	2	5
Nagelkerke R^2	.00	.03	.04

Notes: * $p < .05$. $N = 282$. SES = mean score of father's and mother's education; U-GPA = undergraduate grade point average; GRE-V = GRE Verbal Reasoning; GRE-Q = GRE Quantitative Reasoning; GRE-AW = GRE Analytical Writing. Degree attainment within scheduled time frame was coded 1 = finished in time; 0 = finished later than scheduled time frame.

($\Delta R^2 = .03$, $p < .05$), the three GRE parts added a statistically significant increment to the prediction of G-GPA ($\Delta R^2 = .09$, $p < .001$). Especially GRE-AW was a statistically significant predictor of G-GPA ($\beta = .24$, $p < .05$). Table 3 depicts the results for the logistic regression analysis with degree attainment within scheduled time frame as dependent variable. Compared with the three GRE parts, U-GPA emerged as a valid predictor of degree attainment ($\beta = .45$, $p < .05$). Conducting the same set of analyses based on correlations that were corrected for measurement error in the criterion points to even stronger GRE validities. When correcting for measurement error, the GRE parts explained an additional 15% of the variance in G-GPA above and beyond U-GPA.

In order to test whether the relationship between GRE and G-GPA was only an artifact of SES, we first in-spected the zero-order correlations between GRE scales and parental SES (Table 1). The correlations were not statistically significant. This pattern of results indicated that GRE and SES shared little common variance. The results of the hierarchical regression analysis (Table 2) also revealed that GRE-AW still significantly predicted G-GPA ($\beta = .24$, $p < .01$) when the influence of parental SES was controlled. Accordingly, controlling for SES did not eliminate the predictive power of GRE-AW.

3.1. Supplementary analyses

One may wonder whether GRE validities were similar or differed across study areas. In order to provide some insights on this, we computed zero-order correlations for the two largest subsamples of Psychology and Business students. As can be seen from Table 1, correlations within the Psychology and Business students sample were similar to each other and similar to correlations found in the overall sample. One notable exception was that GRE-Q was significantly related to the degree attainment within scheduled time frame in the Business student sample. In addition, we analyzed whether study area moderates the relationship between the three GRE parts and G-GPA (based on uncorrected correlations) and degree attainment within scheduled time frame. To do so, we formed two dummy variables differentiating between Psychology, Business studies, and the rest (Law, Medicine, and Art and Social Sciences). As can be seen from Tables 4 and 5, none of the interaction terms were significant, indicating that GRE validities equal each other across faculties.

4. Discussion

The present study sets out (a) to investigate the validity of the GRE for study success in a European university considering two different study criteria and (b) to test whether associations found between GRE test scores and study success criteria are independent of parental SES.

Regarding the first goal of the study, our findings provide evidence that the GRE is a valid predictor of graduate study success at a Dutch university. GRE-V, GRE-Q, and GRE-AW measured at the beginning of the master program predicted students' G-GPA at the end of the master program. The finding that none of the other GRE scales predicted the degree attainment within scheduled time frame might be explained by the fact that other, non-ability-related factors may play a larger role in determining whether or not students complete their studies within the designated time frame, for instance financial resources, conscientiousness, motivation, or adaptability.

Table 4. Hierarchical multiple regression analyses predicting G-GPA

	G-GPA	
Predictor	ΔR^2	β
Step 1	.02*	
SES		.15*
Step 2	.03*	
U-GPA		.18*
Step 3	.09***	
GRE-V		.09
GRE-Q		.03
GRE-AW		.24**
Step 4	.08***	
Dummy 1: FPN and other		−.24**
(0) versus SBE (1)		
Dummy 2: FPN and SBE		−.29***
(0) versus Other (1)		
Step 5	.02	
GRE-V × Dummy 1		.09
GRE-V × Dummy 2		.03
GRE-Q × Dummy 1		−.13
GRE-Q × Dummy 2		.11
GRE-AW × Dummy 1		.02
GRE-AW × Dummy 2		−.01
Total R^2	.24	

Notes: * $p < .05$; ** $p < .01$; *** $p < .001$. $n = 236$. G-GPA = graduate grade point average; SES = mean score of father's and mother's education; U-GPA = undergraduate grade point average; GRE-V = GRE Verbal Reasoning; GRE-Q = GRE Quantitative Reasoning; GRE-AW = GRE Analytical Writing; FPN = Faculty of Psychology and Neuroscience; SBE = School of Business and Economics; Other = Faculty of Law, Faculty of Health, Medicine and Life Science, and Faculty of Arts and Social Science.

Overall, the results found in the present study were consistent with previous meta-analyses investigating the predictive validity of the GRE in the US context (Kuncel et al., 2001; Kuncel et al., 2010). Our study results are similar to Kuncel et al.'s (2001) meta-analytically derived average correlations with regard to GRE-V ($r = .21$, present study vs. $r_{obs} = .23$; Kuncel et al., 2001) and GRE-Q ($r = .17$ vs. $r_{obs} = .21$); GRE-AW was not included in the meta-analyses as it did not exist when the studies contributing to the meta-analyses were conducted. Instead of degree attainment, which was defined as the successful completion of a degree program (Kuncel et al., 2001), we focused on degree attainment within the scheduled time frame in the present study. Although degree attainment and degree attainment within scheduled time frame are somewhat different, findings are similar with respect to GRE-Q ($r = .09$ vs. $r_{obs} = .14$). Furthermore, incremental validity evidence obtained in the present study demonstrates that the GRE predicts beyond what could be predicted from U-GPA alone. Especially GRE-AW provides supplemental information about future G-GPA.

In response to previous criticism that associations found between admission test scores and criteria of study success are an artifact of SES (Crosby et al., 2003; Kohn, 2001; Zwick, 2004), a second goal of the present paper was to test whether GRE validity was independent of parental SES. Our findings revealed that (a) GRE test scores were not significantly related to parental SES and (b) that GRE-AW was still significantly related to G-GPA when SES was controlled in a hierarchical

Table 5. Logistic regression analysis predicting degree attainment

Predictor	Model 1	Model 2	Model 3	Model 4	Model 5
SES	.01	−.03	−.04	−.05	−.05
U-GPA		.43*	.45*	.47*	.45*
GRE-V			.01	.01	.00
GRE-Q			.02	.04	.05
GRE-AW			−.23	−.21	−.22
Dummy 1: FPN and other (0) versus SBE (1)				−.85*	−.89*
Dummy 2: FPN and SBE (0) versus other (1)				−.67*	−.66
GRE-V × Dummy 1					.03
GRE-V × Dummy 2					.02
GRE-Q × Dummy 1					−.01
GRE-Q × Dummy 2					−.10
GRE-A × Dummy 1					.12
GRE-A × Dummy 2					−.09
Model fit					
χ^2	0.02	4.82	6.75	13.97	14.38
df	1	2	5	7	13
Nagelkerke R^2	.00	.03	.04	.07	.08

Notes: * $p < .05$. $N = 282$. SES = mean score of father's and mother's education; U-GPA = undergraduate grade point average; GRE-V = GRE Verbal Reasoning; GRE-Q = GRE Quantitative Reasoning; GRE-AW = GRE Analytical Writing; FPN = Faculty of Psychology and Neuroscience; SBE = School of Business and Economics; Other = Faculty of Law, Faculty of Health, Medicine and Life Science, and Faculty of Arts and Social Science. Degree attainment within scheduled time frame was coded 1 = finished in time; 0 = finished later than scheduled time frame.

regression analysis. In order to provide further insights into the role of SES in GRE validities and to compare our findings to previous findings on the role of SES in validities of admission test at the undergraduate level (Sackett et al., 2009), we computed partial correlations (based on uncorrected correlations). Specifically, we partialled out the influence of SES from the correlations between GRE parts and G-GPA and compared these to the uncorrected zero-order correlations reported in Table 1. Controlling for SES, GRE-G-GPA correlations remained virtually the same ($r_{GRE-V-G-GPA} = .21$, $p < .01$; $r_{GRE-Q-G-GPA} = .17$, $p < .01$; $r_{GRE-AW-G-GPA} = .30$, $p < .001$). The latter finding is consistent with Sackett et al.'s (2009) finding from the United States that the SAT–grade relationship was only marginally reduced (the correlation dropped from .47 to .44) when the influence of SES was controlled. The present findings thus replicate Sackett and colleagues' finding at the graduate level. Yet, our study also revealed that GRE scores were unrelated to parental SES and this finding differs from Sackett et al.'s finding that SAT scores were substantially related to SES. Another interesting finding is that U-GPA and G-GPA were related to SES. This indicates that higher SES students tend to get higher grades at both the undergraduate level and the graduate level and that the association between U-GPA and G-GPA may – in part – be caused by parental SES. Provided that this finding can be replicated in future studies using different samples, it suggests that using U-GPA as a selection criterion may be less fair than using a standardized admission test like the GRE that was unrelated to SES in the present study.

Taken together, our results suggest that the GRE is a valid instrument that may meet the urgent need for objective selection instruments in a European context. Besides the predictive and incremental validity evidence obtained in the present study, the GRE meets other important requirements distinctive for a sound selection instrument. First, the GRE is objective as test results are easily interpreted and do not depend on individual evaluations that might be prone to biases. Consequently, the use of the GRE for student selection purposes is robust against legal claims that are often grounded on missing comparability. Furthermore, it is efficient not only from universities' perspectives but also from the perspective of master students. In case European universities would comprehensively agree upon using the GRE for master student selection purposes, students could use their GRE scores for several applications within the same, but also between different countries. Lastly, the GRE could even increase fairness of master students' selection. The present study demonstrated that parental SES has an influence on the relationship between U-GPA and subsequent study success, whereas the relationship between GRE scores and indicators of study success seems to be independent of factors that might be influenced by social class.

4.1. Practical implications

Since 1999, countries within the European Union have worked toward increasing the comparability and compatibility of their higher education programs (Education, Audiovisual and Culture Executive Agency, 2010). One major goal of the Bologna process is to introduce a two-phase system with a bachelors and a master phase in all European countries (of which many have long used a one-phase system). A second goal of the Bologna process is to allow students to change between universities in different regions within European countries and also between European countries as they traverse from undergraduate to graduate education. As students increasingly use these new opportunities, European universities now increasingly need to select students with an undergraduate degree into their graduate programs. Currently, European universities commonly use applicants' U-GPA to select graduate students. U-GPA is easily accessible, and its use is convenient for universities and applicants.

Especially due to the shortcomings of U-GPA, the findings of the present study have important practical implications. In Germany, for example, a debate was going on whether bachelor students graduating from top universities are discriminated against students from less renowned universities when applying for a master program (Dörnfelder, 2010). This criticism is based on the assumption that bachelor students graduating from prestigious universities often get lower grades due to the high requirements compared to students who obtained their bachelor degree from universities with lower standards. Furthermore, changes going along with the Bologna declaration have facilitated students' possibilities to follow a master program which is related, but not necessarily equal, to the study field where the bachelor degree was obtained. As a result, bachelor students from different study fields (e.g., Business and Psychology) compete against each other for the same master programs (e.g., Marketing or Work and Organizational Psychology), although the two programs rely on different grading standards and are therefore difficult to compare. Accordingly, in order to guarantee a fair selection procedure, the GRE is a promising supplement to U-GPA.

4.2. Limitations and future research

The present study is not without limitations. Participants were master students who have already started their master programs. They participated on a voluntary basis and did not practice beforehand. In this sense, the testing situation can be considered a low stakes situation that might have resulted in somewhat lower GRE scores compared with real applicants. This does not negate the validity evidence obtained in the

present study. However, in order to define GRE cut-off scores for a European context, results should be treated with caution.

Another limitation of this study is that we solely collected data at a Dutch university. Although the university is located close to the German and Belgium boarder and characterized by an international student body, future validity studies should be conducted in other European countries as well. Future studies may also benefit from including students from the science, technology, engineering and math (STEM) fields in order to get further information about the predictive validity of the GRE in these fields. It might be especially interesting to investigate whether the inclusion of students from STEM fields will change the predictive validity of GRE-AW as it might be that students more practiced in analytical writing (Humanities and Social Sciences) show markedly different results from those in the STEM disciplines.

A third limitation exists with respect to the sample size of the present study. Researchers have called for large-scale studies and meta-analyses to investigate the characteristics and usefulness of standardized tests (Kuncel & Hezlett, 2010; Sackett, Borneman, & Connelly, 2008). Given that the present study is the first validity study conducted in a European context, we consider it as an initial step toward more comprehensive studies in this field. Large-scale studies would also allow analyzing subgroup differences. Although the adverse impact issue might be more about differences between students originating from different European countries than about ethnic minorities within countries, it should not be neglected.

Finally, a limitation exists with respect to the measurement of SES. In order to measure students' SES, self-ratings of parental education were used, which might be prone to biases. However, a study conducted by Looker (1989) showed that the level of agreement between what students indicate as the highest educational degree of their parents and parents' self-evaluation was high (mother's education: $r = .85$; father's education: $r = .86$). This is in line with Sackett et al. (2009) who noted that students are less likely to provide erroneous information about their parents' education compared to alternative SES measures such as income.

5. Conclusion

The results provide evidence for the usefulness of the GRE in a European context. The GRE allows valid predictions of master students' study success in terms of G-GPA. It facilitates the decision-making process of selection committees who have to decide upon students with various cultural and educational backgrounds.

References

Becker, R., & Kolster, R. (2012). *International student recruitment: Policies and developments in selected countries*. The Hague, the Netherlands: NUFFIC (Netherlands Organisation for International Cooperation in Higher Education). Available at https://www.nuffic.nl/en/library/international-student-recruitment.pdf (accessed January 1, 2012)

Crosby, F. J., Iyer, A., Clayton, S., & Downing, R. A. (2003). Affirmative action: Psychological data and the policy debates. *American Psychologist*, 58, 93–115.

Dörnfelder, A. (2010). Wenn der Abschluss an der Top-Uni zum Nachteil wird. [If graduating at a top-university becomes a disadvantage]. *Handelsblatt*, August 31. Available at http://www.handelsblatt.com/politik/oekonomie/nachrichten/das-master-desaster-wenn-der-abschluss-an-der-top-uni-zum-nachteil-wird-seite-all/3527486-all.html (accessed March 31, 2010)

Education, Audiovisual and Culture Executive Agency. (2010). *Focus on higher education in Europe 2010: The impact of the Bologna process*. Brussels, Belgium: Education, Audiovisual and Culture Executive Agency. Available at http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/122EN.pdf (accessed February 1, 2010)

Goldberg, E. L., & Alliger, G. M. (1992). Assessing the validity of the GRE for students in psychology: A validity generalization approach. *Educational and Psychological Measurement*, 52, 1019–1027.

Hell, B., Trapmann, S., & Schuler, H. (2007). Eine Meta-Analyse der Validität von fachspezifischen Studierfähigkeitstests. [A meta-analytic investigation of subject-specific admission tests in German-speaking countries.]. *Empirische Pädagogik*, 21, 251–270.

Hülsheger, U. R., Maier, G. W., & Stumpp, T. (2007). Validity of general mental ability for the prediction of job performance and training success in Germany: A meta-analysis. *International Journal of Selection and Assessment*, 15, 3–18.

Hunter, J. E., & Schmidt, F. L. (Eds.). (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.) Thousand Oaks, CA: Sage.

Kerckhoff, A. C. (2001). Education and social stratification processes in comparative perspective. *Sociology of Education*, 3–18.

Kohn, A. (2001). Two cheers for an end to the SAT. *Chronicle of Higher Education*, p. B12.

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.

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, 315, 1080–1081.

Kuncel, N. R., & Hezlett, S. A. (2010). Fact and fiction in cognitive ability testing for admissions and hiring decisions. *Current Directions in Psychological Science*, 19, 339–345.

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.

Kuncel, N. R., Wee, S., Serafin, L., & Hezlett, S. A. (2010). The validity of the Graduate Record Examination for master's and doctoral programs: A meta-analytic investigation. *Educational and Psychological Measurement*, 70, 340–352.

Lievens, F., Buyse, T., & Sackett, P. R. (2005). The operational validity of a video-based SJT for medical college admissions: Illustrating the importance of matching predictor and criterion domains. *Journal of Applied Psychology*, 90, 442–452.

Looker, E. D. (1989). Accuracy of proxy reports of parental status characteristics. *Sociology of Education*, 62, 257–276.

Mazzarol, T., & Soutar, G. N. (2002). 'Push-pull' factors influencing international student destination choice. *International Journal of Educational Management*, 16, 82–90.

Morrison, T., & Morrison, M. (1995). A meta-analytic assessment of the predictive validity of the quantitative and verbal components of the graduate record examination with graduate grade point average representing the criterion of graduate success. *Educational and Psychological Measurement*, 55, 309–316.

Nuffic – Nederlandse organisatie voor international samenwerking in het hoger onderwijs. (2006). *Cijfers ontciferd: Notitie aangaande het omzetten van (examen) cijfers: een onderzoek naar de verdeling van cijfers en grades in het onderwijs van Nederland, de Verenigde Staten, Canada en het Verenigd Koninkrijk.* [Grade deciphered: A note regarding the transformation of exam grades: A study on the distribution of grades in higher education in the Netherlands, the United States, Canada, and the United Kingdom]. The Hague, the Netherlands: Author. Available at http:// www.nuffic.nl/root/nuffic/bestanden/documenten/over-de -nuffic/publicaties/factsheet-cijfers-ontcijferd.pdf (accessed January 1, 2006)

Powers, D. E. (2004). Validity of Graduate Record Examinations (GRE) general test scores for admissions to Colleges of Veterinary Medicine. *Journal of Applied Psychology*, 89, 208–219.

Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist*, 63, 215–227.

Sackett, P. R., Kuncel, N. R., Arneson, J. J., Cooper, S. R., & Waters, S. D. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance?. *Psychological Bulletin*, 135, 1–22.

Salgado, J., & Anderson, N. (2002). Cognitive and GMA testing in the European community: Issues and evidence. *Human Performance*, 15, 75–96.

Statistics Netherlands. (2010). Many German students in Dutch higher education. Web magazine. Available at http:// www.cbs.nl/nl-NL/menu/themas/onderwijs/publicaties/ artikelen/archief/2010/2010-3022-wm.html (accessed January 18, 2010)

Zwick, R. (2004). Is the SAT a 'wealth test?' The link between educational achievement and socioeconomic status. In R. Zwick (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (pp. 203–216). New York: RoutledgeFalmer.