Predictive validity of the Slovene Matura exam for academic achievement in humanities and social sciences

Gregor Sočan[*], Maja Krebl, Andreja Špeh and Aneja Kutin

Department of Psychology, Faculty of Arts, University of Ljubljana, Slovenia

Abstract: Matura is a Slovene national examination, which all the students take after successfully completing secondary education. The Matura has two major functions; it is a high school final examination and a selection instrument for University. The goal of the study was to investigate the predictive validity of Matura for predicting academic success in study programmes in the area of humanities and social sciences. Predictive validity was studied both from the traditional correlational perspective and from the multilevel regression perspective. Additionally, we checked for possible differences in predictive validity between study programmes. According to the expectations, the Matura score was a relatively strong and robust predictor of later academic achievement, even after controlling for the high school overall grade. The results support the use of Matura scores in selection of candidates for undergraduate studies in humanities and social sciences.

Keywords: Matura, selection, academic achievement, predictive validity, multilevel modelling

Napovedna veljavnost slovenske mature za študijsko uspešnost v humanistiki in družboslovju

Gregor Sočan[*], Maja Krebl, Andreja Špeh in Aneja Kutin

Oddelek za psihologijo, Filozofska fakulteta, Univerza v Ljubljani

Povzetek: Matura je slovenski državni izpit, ki ga mora po uspešnem zaključku srednješolskega izobraževanja opraviti vsakdo. Ima dve glavni funkciji, in sicer predstavlja zaključni izpit, ki kaže na usvojene standarde znanja v srednji šoli, poleg tega pa služi kot selekcijski instrument za vpis na univerzo. Z raziskavo smo želeli preveriti napovedno veljavnost Mature za napovedovanje študijskega uspeha pri univerzitetnih programih s področja družboslovja in humanistike. Pri tem smo uporabili tako klasični korelacijski pristop kot tudi večnivojski regresijski pristop. Preverili smo tudi morebitne razlike med študijskimi programi. Skladno s pričakovanji se je uspeh na Maturi izkazal za razmeroma močan in robusten napovednik kasnejše študijske uspešnosti, in sicer tudi po nadzoru srednješolskega uspeha. Rezultati podpirajo uporabo Mature v namen selekcije kandidatov za dodiplomske programe družboslovnih in humanističnih smeri.

Ključne besede: matura, selekcija, akademski uspeh, napovedna veljavnost, večnivojsko modeliranje

[*] Naslov/Address: izr. prof. dr. Gregor Sočan, Oddelek za psihologijo, Filozofska fakulteta, Univerza v Ljubljani, Aškerčeva 2, 1000 Ljubljana, Slovenija, e-pošta: gregor.socan@ff.uni-lj.si

Matura is a Slovene national examination which all the students need to take to successfully complete the secondary education. The first official Matura that was obligatory for all high school students was conducted in the school year 1994/1995 and additionally, the Vocational Matura was introduced in parallel to General Matura to replace the internal final examinations in vocational schools. This article is focused on General Matura, which is intended for high-school students.

The General Matura has two major functions. Firstly, it is a high-school final examination. Passing Matura means one has acquired basic standards of knowledge for secondary education and is competent to take any academic course. Having successfully passed Matura is the general condition for applying to University. Secondly, it is a selection instrument for University. If there are more candidates than open positions for a specific study program, performance on Matura usually plays an important role in the admission. The entry requirements differ depending on the specific study. Typically, the weight of the Matura score is 0.6 and the weight of the average grade in the last two years of high school is 0.4 (Budin, 2001).

All the candidates take the exam under the same conditions, that is, at the same time, following the same procedure and rules and with the same evaluation criteria (for details see the annual reports, e.g., Tivadar, 2015). The exam consists of three compulsory and two elective subjects. The compulsory part includes Mother Tongue, Mathematics and a Foreign Language (English, French, German, Italian, Russian or Spanish). For these three subjects one must participate in an oral and written examination. The elective part usually reflects candidate's personal interests. Candidates can choose among (almost) all the other subjects they have encountered during their studies.

Matura is used as a selection instrument for applying to University only when there are more candidates than places available for a specific study. Nevertheless it is still of major importance, because the case of too many applied candidates happens quite often. Therefore it is reasonable to investigate its predictive validity for academic achievement. Cankar (2000) compared the Matura exam and an entrance examination with respect to their predictive validity for the academic success in studying psychology. The entrance examination turned out to be a better predictor of first year average grades in study than Matura. The predictive validity of Matura increased only if grades on optional course of psychology were considered, although it still did not outperform the predictive validity of the entrance examination.

Nevertheless, Matura is used as a major selection tool in Slovenia by law, and the entrance exams have mostly been discontinued. Some studies, mostly art and sports-related studies (see Ministrstvo za šolstvo, znanost in šport, 2015), still require the entrance exams and after passing it Matura plays a role of a second selection tool. Matura stands as a standardized exam which is comparable across all candidates and shows high objectivity and reliability (Bucik, 2001). In combination with the academic success of the 3rd and 4th year of high school, it is usually the only criterion for college

enrolment in case of limited range of accepted students. However, some of aforementioned programmes also require a specific skills or abilities certificate (Pravilnik o razpisu za vpis in izvedbi vpisa v visokem šolstvu [Rules on the call for enrolment and enrolment in higher education], 2016).

Bucik (2001), based on National Examinations Centre's data, argued that Matura scores were lower for students who were enrolled in first year of study twice. The connection of Matura and academic achievement is the highest in first year of study and then it decreases. He also exposed that correlations between Matura scores and academic achievements in programs without admission restrictions were positive, but relatively low. In courses with admission criteria, correlations were still positive, but lower (Bucik, 2001).

Although the education systems differ from country to country, Matura can be roughly compared to other assessments with equal or similar function. In Europe it is comparable among some countries: for example, Czech Republic, Denmark, Ireland, the Netherlands, Norway and Poland all implement a final examination holding the same function as Matura has – showing an accomplishment of knowledge standards and being a selection instrument for University (e.g., Budin, 2001; Bialecki, Johnson, & Thorpe, 2002; Egelund, 2005; Looney, 2006; Strakova & Simonova, 2013; Tveit, 2014). Surprisingly, the published empirical evidence on the predictive validity of final high school examinations in European countries seems to be virtually non-existent.

According to its function, Matura is to a certain extent comparable with other procedures apart from Europe. Kobrin and Patterson (2011) used Scholastic Assessment Test (SAT) scores and grade point average (GPA) as predictors of freshmen's success and found out that the average validity of all different tests was .63. Admissions tests can also be comparable with Matura because they serve as an entry requirement for the higher education process, but it should be noted that the contents differ. Matura covers a broader field of studies and is more related to the high school curriculum than SAT, which covers only three main areas and includes some extracurricular knowledge and specific abilities as well. Admission tests are used in the United States and have good predictive validity for the American students (House & Keeley, 1997). Berry and Sackett (2009) compared the predictive values of SAT scores and high school GPAs. They found out that high school GPAs predicted college grades better than SAT scores did. According to this study, it would make more sense to base the selection of candidates more on the average success in high school than on the Matura results, but it should be taken into account that GPA and Slovene 3rd and 4th year final grades cannot be directly comparable due to a smaller variability of the latter. In United Kingdom, Higher Education Funding Council for England (HEFCE, 2003) concluded that A-level grades were the most important factor in determining the higher education (HE) achievement. Additionally, student's gender, characteristics of the school and the university and subjects studied were also associated with HE achievement.

When the students are selected from a larger group of applicants, the restriction of range may affect the predictive

validity of predictors of academic success. A higher admission criterion means that the selected group of students is more homogenous than the population of applicants (or possible students). Because the variability in the selected group of students is lower than the variability in the population of all candidates, the predictive validity decreases. Most of studies that were trying to estimate the predictive validity of a specific selection criterion hit the problem with including only students that did enter the program, which is only a part of all the students that wanted to study a specific topic (Cankar, 2000). The predictive value of predictors related to the selection procedure is therefore underestimated to some extent. Although "correction formulas" which estimate the correlation in the unrestricted population have been available for a long time (see, for instance, Gulliksen, 1950, ch. 11–13), their practical utility is limited because they are appropriate only for use with relatively simple selection mechanisms. It should be borne in mind, however, that restriction of variability of predictor scores affects only the correlation coefficient, while the regression slope should in principle remain unaffected by selection (see Gulliksen, 1950, p. 131).

Since Matura plays a significant role in the process of applicants selection for academic studies, it is important (from the viewpoints of both fairness and academic efficiency) to establish its predictive value for academic achievements. Previous results on the predictive validity of Slovene Matura (Cankar, 2000) have become outdated and the studies were carried out on Psychology students only. In addition, we did not found any published studies on the predictive value of similar assessments in comparable European countries. Since Matura is a relatively expensive assessment in terms of required financial and organizational resources, it is also important to establish its incremental validity in comparison to the high-school success, which can be obtained at a much lower cost.

The goals of our study were:

1. to estimate the general predictive value of Matura scores for academic achievement across a range of academic study programmes from the area of humanities and social sciences;
2. to determine the incremental predictive validity of Matura over the high-school success;
3. to investigate the differences in predictive validity across various study programmes;
4. to check whether these differences were related to the selection procedure.

With respect to the last point we anticipated a higher predictive validity for programmes with lower or no admission criterion compared to the programmes with low selection ratios.

Two aspects of predictive validity were examined: the correlation coefficient with the academic success and the regression slope for predicting academic success.

We employed two criteria of academic success, namely the average exam grade and a binary variable indicating a successful completion of the studies within the normative four-year period.

Method

Data collection

The data were obtained from the Faculty of Arts of the University of Ljubljana. Faculty of Arts is the biggest faculty in Slovenia with 3504 registered undergraduate students in the 2014/15 academic year, representing 11.5% of all undergraduate students enrolled at the University of Ljubljana (University of Ljubljana, 2013). The faculty consists of 21 departments, which offer 47 programmes (26 dual-subject and 21 single-subject), which are categorised in the areas of social sciences and humanities. The dual-subject courses are a feature of this faculty; in this case, a student simultaneously works on two programmes, each of them comprising of a reduced amount of workload compared to a single-subject programme. After completing both partial studies, a combined degree is conferred. The number of applicants, open positions and admission criteria vary depending on a specific study programme (for details see University of Ljubljana, 2015). The duration of all programmes is three years of organized courses plus an optional additional year in which students can finish all required obligations.

Anonymised student data were obtained from the Student Affair Office at the Faculty of Arts, University of Ljubljana. Initially, our sample included four generations of regular students that were enrolled in an undergraduate study at the Faculty of Arts in the years 2009, 2010, 2011 and 2012. Complete transition to the new Bologna programmes was completed in 2009 and therefore, all students in our sample attended Bologna programmes. Initially, we obtained data for 1985 students. Prior to analyses, we removed the data related to the students who did not take the Slovene General Matura (92 students took the Vocational Matura, 2 students completed the high school before the introduction of Matura, and 60 students completed their secondary education abroad). We also excluded six students with either missing or obviously erroneous data on Matura grades. Finally, we had to exclude 114 students of a programme which consists of separate tracks with different admission criteria, but the students of different tracks could not be differentiated according to the available data. After the exclusion of 251 students[1], our sample finally consisted of data related to 1734 students. In the multilevel predictive analyses (see section on statistical analysis) we only used data of 1481 students who were enrolled between 2009 and 2011. This was necessary because we used the "completion within four years" as a dependent variable.

The following variables were used:

1. Matura grade in points (MG),
2. high-school success (HSS), i.e., achievement in the last year of secondary education,
3. completion within 4 years: a binary variable indicating whether the studies had been concluded within the normative four-year period,

---

[1] Some students met two exclusion criteria, therefore the total number of excluded students was smaller than the sum of the previously stated counts of students meeting particular exclusion criteria.

4. average academic grade (AG): the average of all grades obtained by a student across all years of study,
5. the year of enrolment in the study programme,
6. the study programme of a student,
7. single-subject vs. dual-subject study,
8. the presence of admission restrictions for enrolment,
9. selection ratio (number of accepted / number of applicants; equals 1 if admission was not restricted).

Variables 1-4 were person-level variables, variables 5 and 6 were clustering variables (used to define groups at the second level), and the remaining variables were group level variables (these were characteristics of either the study programme or of the study programme in a particular enrolment year). Variables 3, 7, and 8 were binary variables, and the remaining variables were treated as interval variables.

Table 4 in the Appendix presents the frequencies of students over the study programmes. Note that each student of a dual-subject programme is counted twice (once for each attending programme) and therefore, the total number of the dual-subject students is 1540/2.

The data on admission criteria were obtained from the University website (Higher Education Admissions Office, 2012; Higher Education Admissions Office, 2013).

Statistical analysis

Our data exhibited a multilevel data structure: students were nested within groups, defined by their study program and the academic year of their first enrolment. The intraclass correlation coefficient indicated that about 20% of variance of academic grades could be attributed to the group differences. The predictive equations were therefore analysed using the multilevel modelling (for details see, e.g., Raudenbush & Bryk, 2002). In multilevel analysis, the relation between the dependent variable and its predictors is modelled on two (or more) levels. In our case, the first level was the student level, and the second level was the study group level (i.e., a group of students enrolled in a particular programme or combination of programmes in a particular academic year). The first level regression parameters may be treated as random latent variables on the second level. In our case, we were particularly interested in the question whether the slope for regression of the average academic grade on the Matura grade differs systematically across study groups and whether these differences can be explained by programme characteristics. The particular form of the model used was the cross-classified model, which allows for two crossed (that is, non-nested) classifications on the group-level; in our case, the first and the second study programme.

The null (empty model) for predicting the average academic grade of student i, enrolled in programmes j and k (A_{ijk}; $j = k$ in case of a single-subject study) thus had the following form:

$$\text{Model 0: } A_{ijk} = \pi_{0jk} + e_{ijk} = \theta_0 + b_{00j} + c_{00k} + e_{ijk} \quad (1)$$

where π_{0jk} denotes a random intercept, which can be broken down into the fixed part θ_0 (which is the same for all groups) and random parts b_{00j} and c_{00k}, which differ across

programmes and academic years. Finally, e_{ijk} is the person-level residual.

The predictive random-intercept model had the following form:

$$\text{Model 1: } A_{ijk} = \theta_0 + \theta_1 MG_{ijk} + \theta_2 HSS_{ijk} + b_{00j} + c_{00k} + e_{ijk} \quad (2)$$

where MG and HSS are the Matura and the high-school grades, respectively, and θ_1 and θ_2 are the respective regression slopes. Both the Matura grades and the high-school grades were grand-centered to facilitate the interpretation of the intercepts (which thus become the predicted values for students with average Matura grade and average high-school grade). Then, the slopes for Matura grades were allowed to vary freely across groups, yielding the following model:

$$\text{Model 2: } A_{ijk} = \theta_0 + \theta_1 MG_{ijk} + b_{10j} MG_{ijk} + c_{10k} MG_{ijk} + \\ + \theta_2 HSS_{ijk} + b_{00j} + c_{00k} + e_{ijk} \quad (3)$$

where θ_1 and θ_2 are the fixed (parts of the) regression slopes, which are equal for all groups, while b_{10j} and c_{10k} are the random, group-specific parts of the regression slopes.

Subsequently, the differences between the regression slopes can be explained by group-level variables:

$$\text{Model 3 (level 2): } \pi_{1jk} = \theta_1 + b_{10j} + c_{10k} + \gamma_{11} DS_j + \gamma_{12} SEL_j + \\ + \gamma_{13} SR_j + \gamma_{11} DS_k + \gamma_{12} SEL_k + \gamma_{13} SR_k \quad (4)$$

where π_{1jk} denotes the regression slope for MG, SR is the selection ratio, SEL is a binary variable indicating whether any restrictions were posed for enrolment, and DS is a binary variable indicating whether the programme is dual-subject. We included SEL and SR because we expected a higher predictive validity for programmes with no admission restrictions. We included DS to check for any possible systematic differences between single- and dual-subject programmes. Model 3 is meaningful only in case that Model 2 fits better than Model 1, indicating a significant variance of regression slopes.

In our final model, we additionally allowed the slopes for HSS to vary across groups:

$$\text{Model 4: } A_{ijk} = \theta_0 + \theta_1 MG_{ijk} + b_{10j} MG_{ijk} + c_{10k} MG_{ijk} + \theta_2 HSS_{ijk} + \\ + b_{20j} HSS_{ijk} + c_{20k} HSS_{ijk} + b_{00j} + c_{00k} + e_{ijk} \quad (5)$$

In model 4, regression slopes of both MG and HSS consist of the fixed part θ and the random parts b and c.

Analogous models were stated for predicting the successful completion of the study within four years from beginning. Since the successful completion (C) is a binary variable, we used a multilevel logistic regression, where logit(C) was predicted rather than C itself. The logit transform η_{ijk} of the value of C for student i is defined by:

$$\eta_{ijk} = log \left(\frac{p(C_{ijk} = 1 | \pi_{jk})}{p(C_{ijk} = 0 | \pi_{jk})} \right) \quad (6)$$

Therefore, the predicted variable was the log-odds for successful completion.

For the multilevel modelling we used the HLM 7 software (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011). We used the default estimation methods (full maximum likelihood for predicting the academic grade, and penalised quasi-likelihood for predicting the successful conclusion of the study).

To check whether the size of predictive validity coefficients (i.e., the correlations between Matura grades and academic success) varied across programmes (possibly in relation to the selection ratio), we first checked whether the distribution of validity coefficients was more dispersed as could be expected on base of random sampling. Unfortunately, correlations cannot be modelled by means of multilevel modelling. Instead, we used a permutation test. We proceeded as follows.

The null hypotheses we tested stated that:

a. $\sigma^2_{\rho(AM)} = 0$, that is, all differences between within-group correlations across groups can be attributed to chance;
b. the actual distribution of correlation coefficients ρ_{AM} is the same as the distribution that arises due to sampling error.

1. We computed the Pearson correlation coefficient between both grades in each subgroup to estimate the actual distribution of correlations. Only groups with at least five students were included (the total sample size was thus reduced to 1199).
2. We standardized grades within groups to prevent any confounding effects due to group differences in average grades.
3. We randomly permuted the pairs of grades and assigned them to groups of the same sizes as were the actual groups. As with the actual data, correlations were computed in each group. This step was repeated 10.000 times.
4. Finally, we tested the first H0 by determining the percentile rank of the actual variance of correlation coefficients with the distribution of variances obtained by the permutation test. We tested the second H0 by means of the two-sample Kolmogorov-Smirnov test.

We performed the permutation test by means of the R 3.1.3. software. (R Core Team, 2015). For all hypotheses testing, we set the alpha error rate at 5%.

Results

Descriptive statistics

Table 1 presents descriptive statistics for the non-binary person-level variables. Note that the possible value ranges were [6, 10] for academic grade, [10, 34] for Matura grade and [2, 5] for the high-school grade. All distributions were close to symmetric and exhibit some negative kurtosis, which should be expected because of the limited range of possible values. In our sample, 73.1% of the students successfully finished their study not later than 4 years after enrolment.

Table 1. *Descriptive statistics for person-level variables*

	AG	MG	HSS
M	8.16	21.83	4.00
Mdn	8.06	22	4
*Q*1	7.69	18	3
*Q*3	8.59	26	5
SD	0.63	5.33	0.82
Skewness	0.47	0.12	−0.33
Kurtosis	−0.31	−0.60	−0.70
Min.	6.72	10	2
Max.	10.00	34	5

Note. $N = 1481$. M = mean, Mdn = median, Q = quartile, AG = average academic grade, MG = Matura grade, HSS = high-school grade.

Multilevel models

In the first multilevel analysis we predicted the average academic grade. The main results are presented in Table 2. In the empty model (model 0), the intraclass correlation coefficient was .20, indicating a considerable amount of variance at the group level. The multilevel modelling was therefore necessary.

In model 1, we added the Matura grade (MG) and the high-school success (HSS) as predictors at the student level. The estimated academic grade for a student with average values of both MG and HSS was 7.6. Both predictors had statistically significant slopes. A point increase in MG (roughly corresponding to 1/5 of standard deviation) was associated with a grade increase of 0.07, while a unit increase in HSS was associated with a grade increase of 0.15. However, the MG had a higher predictive power as HSS. The inclusion of MG only would reduce the unexplained student-level variance for 37%, while the inclusion of HSS would result in a 22% reduction. The inclusion of both predictors resulted in 39% reduction of unexplained variance.

In model 2, we allowed the regression slopes for MG to vary across groups. The deviance test did not show a significant improvement in model fit ($p = .32$), and the variance components for MG slopes were very small. Therefore, we found no evidence of different regression slope for Matura grades in predicting academic grades in different study programmes in different academic years. For illustration purposes, we nevertheless predicted the slope residuals by three group characteristics (presence of selection, selection ratio, and one- vs. dual-subject programme) in model 3. As expected, all three coefficients were close to zero and not statistically significant. In model 4, we allowed the slopes for HSS to vary across groups. Again, the deviance test did not show a significant improvement of fit ($p = .99$). Therefore, a unit increase in HSS is not associated with a statistically different increase in academic grade in different study groups.

Table 2. *Fixed and random effects for models of academic grade*

	Model 0		Model 1		Model 2		Model 3		Model 4	
	Fixed effects									
Student level										
Intercept	8.19*	(.03)	7.66*	(.09)	7.64*	(.09)	7.64*	(.09)	7.65	(.10)
MG			0.07*	(.00)	0.07*	(.00)	0.07*	(.03)	0.07	(.03)
HSS			0.15*	(.02)	0.15*	(.02)	0.15*	(.02)	0.15	(.02)
Group level (slope for MG)										
SR							0.00	(.03)	0.00	(.03)
DS							0.00	(.01)	0.00	(.01)
SEL							−0.01	(.01)	−0.01	(.01)
	Random effects									
	VC	*SD*	VC	*SD*	VC	*SD*	VC	*SD*	VC	*SD*
Student	0.3240	0.57	0.1967	0.44	0.194	0.44	0.194	0.44	0.193	0.44
Intercept1	0.0351	0.19	0.0739	0.27	0.077	0.28	0.076	0.28	0.110	0.33
Intercept2	0.0481	0.22	0.0272	0.16	0.026	0.16	0.027	0.16	0.022	0.15
Slope (MG) 1					0.000	0.01	0.000	0.01	0.000	0.01
Slope (MG) 2					0.000	0.00	0.027	0.16	0.000	0.00
Slope (HSS) 1									0.001	0.02
Slope (HSS) 2									0.002	0.05
deviance	2568.3		1928.6		1923.9		1922.8		1922.0	
n par.	4		6		10		13		19	
$p(M_i$ vs. $M_{i-1})$					.32				.99	

Note. For fixed effects, coefficients and parenthesised standard errors are presented.
VC = variance component, *SD* = standard deviation, *n* par. = number of estimated parameters,
$p(M_i$ vs. $M_{i-1}) = p$ value related to the test of two models with nested random effects, SR = selection ratio,
DS = dual-subject study, SEL = binary variable indicating whether the candidates were selected (SEL = 1) or not (SEL = 0).
*$p < .05$ (for tests of fixed effects)

In the second analysis, the successful conclusion of the study in four years was predicted. Only the results for models 1 and 2 are presented in Table 3. In model 1 (model with predictors included and random intercepts), both MG and HSS were statistically significant predictors: the odds ratios were 1.101 for MG and 1.537 for HSS, respectively. Therefore, a point increase of the Matura grade increased the odds of successfully finishing the study within four years for about 10%. The logistic regression coefficients and the associated odds-ratios remained practically unchanged in model 2 when the slopes were free to vary across groups. There were no statistically significant random effects[2] in either model 1 or model 2. This means that neither the log-odds for a successful conclusion nor the regression slopes for predicting these log-odds on Matura grades differed significantly across groups. Therefore, we again found no evidence of different predictive validity of Matura grades over the programmes.

Variability of predictive validity coefficients across groups

Despite the fact that no evidence for differences in regression slopes for Matura grades was found, the predictive validity coefficients (i.e., correlations between Matura grades and academic grades) could still differ across groups because of the differences in variability of either variable across groups. In the second part of our study we therefore checked whether the correlation coefficient between Matura grades and academic grades depends on the programme or programme combination, respectively. Seemingly, the distribution of the correlation coefficients was quite dispersed: The coefficients ranged from −.33 to .98 with a mean[3] of .61, a median of .62, and a standard deviation of .26. However, many of the groups were quite small, so it was necessary to check whether these differences could be attributed to the sampling error. As

---

[2] In cross-classified models with a binary dependent variable, the deviance test for models differing in random effects is not available. We therefore need to rely on tests for specific variance components.

[3] We computed the mean via the weighted average of the Fisher's *z* values.

Table 3. *Fixed and random effects for models of a successful conclusion*

	Fixed effects					
	Coefficient	*SE*	*t*	*p*	*OR*	95% CI (*OR*)
Model 1						
Intercept	1.422	0.121	11.77	< .001	4.15	(3.27, 5.26)
MG	0.097	0.020	4.85	< .001	1.10	(1.06, 1.14)
HSS	0.430	0.121	3.56	< .001	1.54	(1.21, 1.95)
Model 2						
Intercept	1.393	0.117	11.90	< .001	4.03	(3.20, 5.07)
MG	0.096	0.023	4.18	< .001	1.10	(1.05, 1.15)
HSS	0.435	0.121	3.59	< .001	1.54	(1.22, 1.96)

	Random effects					
	SD	VC	*df*	χ^2	*p*	
Model 1						
Intercept 1	0.771	0.595	122	111.9	> .500	
Intercept 2	0.349	0.122	123	120.9	> .500	
Model 2						
Intercept 1	0.693	0.480	100	76.8	> .500	
Intercept 2	0.338	0.114	104	103.8	> .500	
Slope (MG) 1	0.090	0.008	100	45.3	> .500	
Slope (MG) 2	0.016	0.000	104	91.2	> .500	

Note. SE = standard error, *OR* = odds ratio. See also notes to the previous tables.

previously explained, we used a permutation test to test the null hypothesis that the variance of correlation coefficients across groups is zero. The result was not significant (*p* = .86) indicating that the obtained differences between the validity coefficients could easily arise by chance alone. In fact, the expected standard deviation of the distribution of correlation coefficients, as estimated by the permutation test, was .29;

Figure 1. Empirical and expected values of predictive validity coefficients. Solid line: empirical correlations; dashed line: expected correlations under H0.

therefore, the actual correlations differed across groups slightly less than expected.

We compared the empirical distribution of validity coefficients with the expected distribution under the null hypothesis. The two-sample Kolmogorov-Smirnov test did not show a statistically significant difference between distributions (*D* = 0.101, *p* = .297). Figure 1 presents probability densities based on both distributions. A remarkable similarity of the distributions can be seen clearly. Therefore, the distribution of the validity coefficients across groups could arise by chance and should not be taken as evidence of different levels of validity of Matura grades for predicting the average academic grade.

The distribution of actual correlations was markedly negatively skewed (skewness = –1.38), therefore the median correlation coefficient (r_{Mdn} = .62) is a more representative measure of the central tendency as the mean. However, both the median and the average correlation imply a very similar amount of explained variance as found with multilevel models.

Discussion

The goal of our study was to verify the value of General Matura as a selection instrument for University study and to investigate possible differences in predictive validity across different study programmes. We have confirmed that Matura is an important predictor for academic achievement, explaining about 37% variance of mean academic grades. When controlling for the overall success in the final year
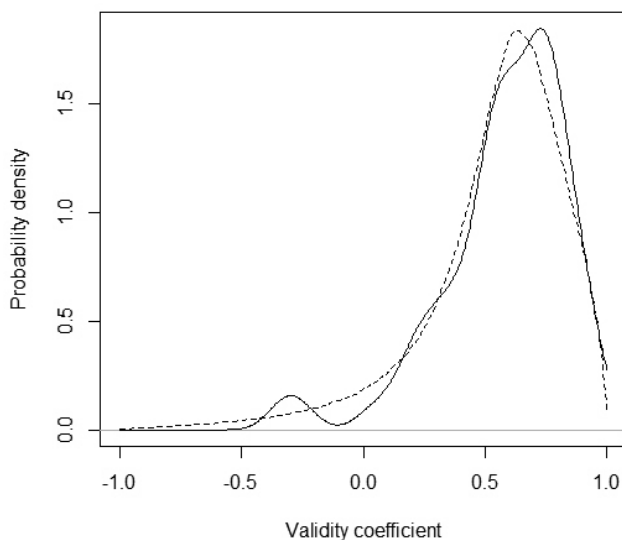
of the high school, a point increase in the Matura score implied an increase of 0.07 of the average academic grade and a 10% increase of the odds ratio for a timely graduation. In comparison to a similar study by Cankar (2000), we have confirmed the predictive validity of Matura grades as a selection tool for academic studies. Additionally, we have generalised this finding to a broad (although not exhaustive) scope of social science and humanities studies. We have also added the regression perspective to the correlational perspective prevalent in studies of predictive validity. Our results indicate a higher predictive validity than Cankar's study. This may be a consequence of the improvement of the Matura tests in the meantime period. However, Cankar's sample size was much smaller than ours, so the differences may at least partly be due to sampling fluctuations.

Our results might also be interesting from an international perspective, as they augment the scarce empirical evidence on the predictive validity of similar examinations across Europe. In the United States of America (USA), as aforementioned, SAT and GPA stand as the main selection procedures and were proved to be valid predictors of success (House & Keeley, 1997; Korbin & Patterson, 2011), which is comparable with results of our study. We discovered that Matura and average grade in the final year of high school both predicted academic success. We should note that the SAT and Matura are not completely equivalent; the contents of Matura are broader in scope and more closely connected to the high school curriculum, while the SAT tests cover only three main areas and exhibit elements of an ability test. In contrast to the USA based results of Berry and Sackett (2009) who found GPA to be a better predictor of academic success as SAT, in our study Matura was a stronger predictor than the high-school success. We can only speculate about the causes of this disparity. It may be a consequence of a broader content of Matura in comparison to SAT. On the other hand, the predictive power of the high school success was probably somewhat underestimated in our study, because we only had data for the final year of the high school, while the success in the final two years is actually used for selection purposes.

One of the reasons for lower predictability of HSS might also lie in grade inflation. Grade inflation is a phenomenon where higher average school grades throughout the years do not correspond to increased knowledge of students (Rosovsky & Hartley, 2002), which is also occurring in the Slovene education system (e.g., Zupanc & Bren, 2010). Because the range of possible grades is limited, higher grades eventually imply less variability among them, and a smaller range of scores may lower the predictive power of HSS.

Bucik (2001) expected that Matura grades would have a lower predictive validity for the academic performance of students enrolled to study courses with higher selection ratio, compared to study courses without admission criteria. Our results did not confirm this expectation: the distribution of validity coefficients across study programs was not statistically different to the distribution that could have arisen by chance. Although this finding is not a proof of equality of validity coefficients (because the null hypothesis cannot be proven in general), it is an evidence of relative robust predictive properties of Matura scores for the academic success, especially taken together with the finding that the regression slopes did not differ significantly across study programmes, regardless of which criterion of academic success was predicted.

We should note some limitations of our study. The most obvious one is the limited scope of study programmes. Although the Faculty of Arts at the University of Ljubljana offers a large number of study programmes in the area of social sciences and humanities, it still does not represent some studies in this area such as economy, journalism and law. Additionally, the proportion of linguistic and philological programmes is relatively large. Future studies should address the question of the predictive validity of Matura for science and technology programmes as well. Due to administrative reasons, the scope of data at our disposal was limited. We did not have access to the components of the Matura grade (i.e., the grades for each subject separately) and to high-school grades in the third year. With such data, more refined questions could be studied, for instance, the possibility of reliably optimising the weights for the elements of the composite of Matura exams and high-school success in order to maximise the predictive validity for a particular study programme.

With regard to the statistical analysis, one might object treating the grades as interval variables. However, these variables are treated as such in all administrative procedures (including the selection of prospective students), and we aimed to investigate the predictive validity of Matura in real rather than ideal conditions. Again, constructing more refined predictive and selection models that would be perfectly suited to the nature of the data, remains a task for future research. Also, the high school grades are probably not perfectly equivalent across schools (or even across teachers), because teachers' "calibration of grading" can never be completely standardised in practice. In our study, we deliberately disregarded this effect because it is disregarded in the selection procedures as well; but it might be important in studies attempting to optimise the prediction of academic success regardless of the administrative regulations. Further, our research problem was not a typical case for the cross-classified multilevel model, because both classifications were not qualitatively different. However, in our opinion this model is still the most appropriate available model for a multilevel analysis of the data at hand.

To conclude, our study found a relatively high predictive validity of the Slovene General Matura for predicting academic success in humanities and social sciences. The level of predictive validity seems to be quite robust to factors like the differences between programmes, restriction of range due to selection etc. Still, more research is needed to generalise these findings to other study disciplines and to investigate the possibility of constructing more efficient selection formulas.

Acknowledgement

References

Berry, C. M., & Sackett, P. R. (2009). Individual differences in course choice result in underestimation of the validity of college admissions systems. *Psychological Science, 20*(7), 822–830.

Bialecki, E., Johnson, S., & Thorpe, G. (2002). Preparing for national monitoring in Poland. *Assessment in Education: Principles, Policy & Practice, 9*(2), 221–236.

Bucik, V. (2001). Napovedna veljavnost slovenske mature [Predictive validity of the Slovene Matura]. *Horizons of Psychology, 10*(3), 75–87.

Budin, J. (2001). *Zasnova slovenske mature in mednarodne primerjave: Tretji strokovni posvet o maturi: Zbornik povzetkov [The Design of Slovene Matura and International Comparisons: The Third Expert Consultation on the Matura: Book of Abstracts].* Ljubljana, Slovenia: Državni izpitni center.

Cankar, G. (2000). Napovedna veljavnost mature za študij psihologije [Predictive validity of Matura for the study of psychology]. *Horizons of Psychology, 9*(1), 59–68.

Egelund, N. (2005). Profiles of education assessment systems worldwide: Educational assessment in Danish schools. *Assessment in Education: Principles, Policy & Practice, 12*(2), 203–212.

Gulliksen, H. (1950). *Theory of Mental Tests.* New York, NY, USA: Wiley.

Higher Education Funding Council for England (2003, July). *Schooling Effects on Higher Education Achievement* (Issue paper No. 32). Retrieved from /http://www.hefce.ac.uk/pubs/hefce/2003/03_32/03_32.pdf

Higher Education Admissions Office. (2012). *Predstavitev podatkov po študijskih programih [Presentation of Data by Study Programme].* Retrieved from http://www.vpis.uni-lj.si/Analiza%20arhiv/2011_2012/pdf/18.pdf

Higher Education Admissions Office. (2013). *Predstavitev podatkov po študijskih programih [Presentation of Data by Study Programme].* Retrieved from http://www.vpis.uni-lj.si/Analiza%20arhiv/2012_2013/pdf/18.pdf

House, J. D., & Keeley, E. D. (1997). Predictive validity of college admissions test scores for American Indian students. *The Journal of Psychology, 131*(5), 572–572.

Kobrin, J. L., & Patterson, B. F. (2011). Contextual factors associated with the validity of SAT scores and high school GPA for predicting first-year college grades. *Educational Assessment, 16*(4), 207–226.

Looney, A. (2006). Profiles of education assessment systems worldwide: Assessment in the Republic of Ireland. *Assessment in Education: Principles, Policy & Practice, 13*(3), 345–353.

R Core Team. (2015). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods.* Thousand Oaks, CA, USA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & du Toit, M. (2011). *HLM 7: Hierarchical Linear and Nonlinear Modeling.* Chicago, IL, USA: Scientific Software International.

Ministrstvo za šolstvo, znanost in šport RS. (2015). *Razpis za vpis v dodiplomske in enovite magistrske študijske programe v študijskem letu 2015/2016 [Rules on the call for enrolment in undergraduate and uniform master's degree programmes in the academic year 2015/2016].* Retrieved from http://www.mizs.gov.si/fileadmin/mizs.gov.si/pageuploads/Visoko_solstvo/Razpis_2015/Dodiplomski/Javni/razpis_2015_2016_ena_datoteka.pdf

Rosovsky, H., & Hartley, M. (2002). *Evaluation and the Academy: Are We Doing the Right Thing? Grade Inflation and Letter of Reccomendation.* Cambridge, MA, USA: American Academy of Arts & Science.

Pravilnik o razpisu za vpis in izvedbi vpisa v visokem šolstvu [Rules on the call for enrolment and enrolment in higher education], Uradni list Republike Slovenije [Official journal of the Republic of Slovenia], No. 32 (21.1.2016). Retrieved from http://www.uradni-list.si/1/objava.jsp?sop=2016-01-0199

Strakova, J., & Simonova, J. (2013). Profiles of education assessment systems worldwide: Assessment in the school systems of the Czech Republic. *Assessment in Education: Principles, Policy & Practice, 20*(4), 470–490.

Tivadar, H. (Ed.). (2015). *Splošna matura 2015: Letno poročilo [General Matura: Annual Report].* Ljubljana, Slovenia: Državni izpitni center.

Tveit, S. (2014). Profiles of education assessment systems worldwide: Educational assessment in Norway. *Assessment in Education: Principles, Policy & Practice, 21*(2), 221–237.

University of Ljubljana. (2013). *Data on the Number of Attending Students (1st, 2nd, and 3rd degrees).* Retrieved from https://www.uni-lj.si/university/university_in_numbers/

University of Ljubljana. (2015). *Letno poročilo filozofske fakultete 2014 [Annual Report of the Faculty of Arts 2014].* Retrieved from http://www.ff.uni-lj.si/Portals/0/Dokumenti/Porocila/Kakovost/Poslovno%20poro%C4%8Dilo%20s%20poro%C4%8Dilom%20o%20kakvosti%20FF%20UL%202014.pdf

Zupanc, D., & Bren, M. (2010). Inflacija pri internem ocenjevanju v Sloveniji [Inflation in the internal grading in Slovenia]. *Sodobna pedagogika, 61*(3), 208–228.

Appendix

Table 4. *Number of students by programme*

Programme	SS	DS
Archaeology	40	
Art History	10	39
Bohemistics		16
Classical and Humanistic Studies	6	
Comparative Linguistics		8
Comparative Literature and Literary Theory	30	154
Comparative Slavic Linguistics		4
East Asian Cultures		12
English Studies	61	161
Ethnology and Cultural Anthropology	80	23
French and Romance Studies	24	
French Studies		70
General Linguistics		12
Geography	116	122
Germanic Studies	43	67
Greek Language, Literature, and Culture		4
History	26	130
Italian Language and Literature		37
Japonology	23	19
Latin Language, Literature, and Culture		7
Library and Information Science	103	
Musicology	32	
Pedagogy and Andragogy	70	83
Philosophy	8	81
Polish Studies		12
Psychology	197	
Russian Studies		67
Sinology	15	
Slovak Studies		5
Slovene Studies	47	158
Sociology		106
Sociology of Culture	33	
South Slavic Studies		34
Spanish Language and Literature		97
Theological studies		12
Total	964	1540

Note. SS = single-subject study; DS = dual-subject study. Total sample size = 964 + 1540/2 = 1734.