Cankar, Gasper

Governing the transition to higher levels of education and differences between achievement and school grades by gender

Journal for educational research online 8 (2016) 2, S. 60-86

Journal for Educational Research Online
Journal für Bildungsforschung Online
Volume 8 (2016), No. 2, 60–86

Gašper Cankar

Governing the transition to higher levels of education and differences between achievement and school grades by gender

Abstract

The author will explore the performance of boys and girls in external examinations in Slovenia at the beginning of upper secondary and tertiary education. These are critical points in students' educational career at which he/she has to choose a school/university. Since both transitions are managed centrally by appropriate authorities, this is also a question of Educational Governance. Transitions between levels of education should, above all, assure fairness in selection procedures.

At the point of transition to upper secondary schools we will explore differences between students' achievements in various school subjects tested at the national assessment of knowledge (NA), and their school grades by gender. Since only school grades are used as admission criteria to upper secondary schools, this comparison of school grades with external and more objective measure of students' achievement will show possible bias. In Slovenia admission to tertiary education usually consists of (externally assessed) Matura results and school grades in the last two years of upper secondary school. The Author will compare the effects of both most commonly used measures of academic achievement on admission in view of gender differences. Study courses where selection procedure was actually applied will be of specific interest since they can show signs of (un)fairness. Results show signs of bias and build case for better Educational Governance.

Keywords

Matura; Gender differences; External examinations; Achievement scores, Bias

Gašper Cankar, PhD, National Examinations Centre, Kajuhova 32 U, SI–1000 Ljubljana, Slovenia
e-mail: gasper.cankar@guest.arnes.si

Steuerung des Übergangs zu höherer Bildung und geschlechtsspezifische Unterschiede zwischen Leistung und Schulnoten

Zusammenfassung

Der Beitrag untersucht die Leistungen von Jungen und Mädchen in zentralen Prüfungen zu Beginn der Sekundarstufe II und der Hochschulbildung in Slowenien. Bei den Übergängen handelt es sich um kritische Zeitpunkte in der Bildungskarriere von Schülerinnen und Schülern, zu welchen sie eine Schulform bzw. Universität wählen müssen. Da beide Übergänge zentral durch zuständige Behörden gesteuert werden, stellt dies ein Feld der Educational Governance dar. In erster Linie sollen Bildungsübergänge Gerechtigkeit in den Selektionsprozessen gewährleisten.

Am Übergang zur Sekundarstufe II werden – differenziert nach Geschlecht – Unterschiede zwischen den durch das National Assessment of Knowledge (NA) getesteten Leistungen von Schülerinnen und Schülern in verschiedenen Schulfächern einerseits und deren Schulnoten andererseits untersucht. Da allein die Schulnoten als Zugangskriterium zur Sekundarstufe II dienen, kann der Vergleich zwischen Schulnoten und zentralen, objektiveren Leistungsmessungen mögliche Verzerrungen aufzeigen. Die Zugangsvoraussetzung zur Hochschulbildung ergibt sich in Slowenien üblicherweise aus den Ergebnissen der (zentralen) Abschlussprüfung und den Schulnoten der letzten zwei Jahre der Sekundarstufe II. Der Autor vergleicht die Effekte dieser beiden meisteingesetzten Schulleistungsmessungen auf die Zulassung im Hinblick auf Geschlechterunterschiede. Studiengänge, für die Auswahlverfahren stattfinden, sind von besonderem Interesse, weil sie Hinweise auf (Un-)Gerechtigkeit liefern können. Die Ergebnisse deuten auf Verzerrungen hin und liefern Hinweise für die Verbesserung der Steuerung im Bildungssystem.

Schlagworte

Hochschulreife; Geschlechterunterschiede; Zentrale Prüfungen; Leistungsergebnisse, Verzerrungseffekte

1. Introduction

Modern educational systems should employ evidence-based approaches in order to create informed decisions about any important topic. To work in the best interest of its citizens, the government must provide rationale and supporting evidence for any measures that have vast impact on individuals. Lack to do so could be seen as sub-optimal governance (Slavin, 2002).

1.1 Equity

In Slovenia research about equity in education is still rare. There was some improvement when PISA 2006 results showed problematic proportion of between schools variance in student performance (OECD, 2007, Table 4.1a) where Slovenia was one of three countries with greatest between-school variance. Large portion of between-school variance was also explained by PISA index of economic, social, and cultural status of students and schools – schools varied greatly in their intake of students by performance and by socio-economic characteristics. Although secondary analyses (Gaber, Cankar, Marjanovič Umek, & Tašner, 2012) showed that PISA results might be misleading if context is not taken into account, the results still had a sobering effect that things we always took for granted (in this case equity of educational opportunities) might not be true.

Questions regarding equity and fairness of educational system for all its participants are important since they remind us what we neglected, forgot, or were simply unaware of. Procedures that are in a Slovenian educational system well established and are in some cases (i.e. General Matura) already parts of our educational tradition are being questioned with new rigor in search of supporting evidence for equity or lack thereof. In this article we will explore the governance of transitions to higher levels of education and try to search for evidence of (un)fairness in educational system. We will explore differences between boys and girls in their school grades and achievement scores but we will not focus on the differences *per se* but in the question of whether the selection procedures that use those data can still be declared *fair* with regard to any differences found. This research is more focused on the differences in bias, shown by different types of criteria, used in selection procedures, than with differences between results for boys and girls. In this research equity (or fairness) in the selection procedures could be operationalized in a question: Is selection procedure still fair if some measures of student's achievement show significantly greater differences between boys and girls than other while all measures should provide information on (same) student achievement?

1.2 Educational Governance

As Coward (2010) points out Educational Governance is about "exercising authority, control and direction" (p. 711) and the term describes "diverse approaches to designing, funding and managing education to benefit learners, organizations and wider society" (p. 710).

Educational Governance should rely on the feedback from the educational system which should evaluate effects of high stakes decisions. To ensure valid and systematic approach to educational improvement those decisions should be based on evidence and driven by appropriate data (Slavin, 2002). De Coster, Forsthuber, Oberheidt, Parveva, & Glass (2008) note that profound changes in European educational systems have led to reviews of established governance struc-

tures. Those are often supported through international integrations either within European Union or globally within Organization for Economic Co-operation and Development (OECD), which are providing platforms where countries can compare and reflect on differences in educational systems. Calls for greater efficiency and equity in educational systems can be often traced to documents and books, published by OECD (2003; Fazekas & Burns, 2012). High standards of achievement for all students and equity in admission procedures are seen as important aspects of modern higher education (Hénard & Mitterle, 2010). In this regard transitions between levels of education can be seen as important points of high leverage for the stability and effectiveness of the educational system as a whole.

Since the beginning of the Bologna process Europe is becoming more and more relevant level for governance (Krucken, 2011). Maybe admission to tertiary education will eventually become uniformed across Europe and some common guidelines will be established to assure equity and fairness, but this is not very likely in near future. Until then, transitions between educational levels in Slovenia are centrally managed through ministry itself or its appointed body. This makes a selection procedure also an important point of Educational Governance.

1.3 First nine years of schooling in Slovenia

In Slovenia, children enter compulsory nine year schooling at the age of six. Typical elementary school will have all nine grades and for the first nine years of schooling children go to one institution that's closest to their home – a comprehensive school. After they finish ninth grade, they have to choose an upper secondary school to go to. They can choose among gymnasiums that prepare students for tertiary education or technical and vocational educational tracks that offer profession/vocation as shown in Table 1.

Table 1: Upper secondary education by educational track (Eurydice, 2014)

	Duration	External examination
General gymnasium	4 years	General Matura
Professional gymnasium	4 years	General Matura
Upper secondary vocational schools	4 years	Vocational Matura
3-year-vocational schools	3 years	–
2-year-vocational schools	2 years	–

1.4 Selection procedure from comprehensive to upper secondary schools

Every year a ministry publishes the list of available places in the upper secondary schools. This is the document that defines the number of available places for each school and each educational program within the school. Students then in March send their applications to upper secondary schools. After applications are processed in the middle of April, the number of applications per schools is published and after that there is some limited time to transfer the applications to another school (that still has more places that applications). In May the names of the schools where there are still too many applicants are published – only on those schools the selection procedure will take place. Selection procedure is centrally implemented by Ministry.

Prior to 2006, the selection criteria was combined – one criteria were grades from last three years in school, other criteria was result on external examination in Slovene language and mathematics in Grade 9. There were also additional points for high achievements on national competitions in mathematics, literature, chemistry, etc.

1.5 Implementation of national assessment

In 2006 the national assessment was implemented on whole population. This replaced the external examinations in Slovene language and mathematics that existed before and were used in selection procedure for upper secondary schools. At the same time the ministry decided to change the function of national assessment from high stakes selection instrument to a low-stakes feedback information. Since 2006 on, therefore only school grades are used for selection procedure.

1.6 Inflation of school grades

As Willingham and Cole (1997) point out, fair assessment is very important and standards for such tests are hard to achieve. However, school grades, when used in same high stakes decisions as test results, should undergo similar scrutiny. Since school grades, as Zupanc and Bren (2010) noted, in the last years in Slovenia became severely inflated, schools have a problem. This is furthermore aggravated by the fact that some school subjects (i.e., physical education, music, arts, ...) revised their grading system and got numerical grades identical to those of other subjects, and are consequently used in the calculation of average school grade. Those subjects traditionally award high grades to most students and consequently decrease the variability of the grades used for selection.

Schools with 'numerus clausus' are therefore receiving applications from groups of students with almost maximal number of points and a selection criterion is not

very discriminative. National assessment is in the meantime becoming accepted as a useful and valuable tool that provides helpful feedback information for individuals and on system level. It provides accurate and objective measure of individual's knowledge since the tests are comprehensive, curriculum based, and uniform for the whole population and they discriminate also on the top of the scales where students applying for the schools that require selection are. The problem of using national assessment for selection procedure now is conceptual – if used for selection purposes, the instrument loses its low stakes formative function and will be implicitly or explicitly redesigned to reflect this new high stakes function. This would thwart validity of assessment as feedback instrument and reduce its function to an admission test.

As a compromise, there is a limited possibility to use points from national assessment tests in selection procedure only if there are many students with same result on other criteria around the cut score and they couldn't be differentiated otherwise. This is enacted rarely and for a handful of students each year only.

Although there are small changes each year, the selection procedure remains a point of discontent for parents, students, and teachers alike. Many ideas and propositions were entertained but so far none was accepted.

1.7 Admission to tertiary education

Large portion of population will continue their academic career path into tertiary education. Admissions office that provides national services for admission to tertiary education reports in its yearbook (VPIS UL, 2014) that over 80 % of students in the last year of upper secondary schools will enlist into study courses (see Table 2). This proportion means that most students finishing General and Vocational Matura continue their study on tertiary level of education. Seen from the viewpoint of the individual, the transition into tertiary education is even more important than transition to upper secondary level. Stakes are higher since it directly influences choice of profession, employment options, and standing on labor market. It is important to note that admission process in Slovenia is centrally organized and most study courses have very similar criteria if the number of students exceeds the number of available places. The effects of established criteria aren't well researched mostly due to lack of available databases that would join relevant data from upper secondary schools, universities, and labor market. The selection process at the transition into tertiary education is therefore a question of Educational Governance and we can ask ourselves about the fairness and validity of existing procedure.

Table 2: Proportion of students in the last year of upper secondary education applying for tertiary education (VPIS UL, 2014)

Study year	No. of applications from students in last year	No. of all students in last year	Proportion of students sending application
2009/2010	17,051	20,466	83.3 %
2010/2011	16,501	19,915	82.9 %
2011/2012	15,546	18,502	84.0 %
2012/2013	14,780	17,868	82.7 %
2013/2014	14,322	17,626	81.3 %

1.8 Admission procedure

In February each year, the number of available places in each study course is being published by the Ministry of education, science, and sports (MIZŠ, 2014). Each study course must specify the criteria for selection if there would be too many candidates. Selection criteria themselves don't change often, since they are determined when accreditation for the study course is being awarded. Most of the study courses have very simple criteria – 60 % of points are contributed by success on General or Vocational Matura, while 40 % of points by the final grade[1] in the last two years of upper secondary school. Other large group of study courses has same selection criteria but in reversed proportions (40/60). Only few programs are more specific or test additional talents like arts, music, and architecture. For example, one of most complex set of selection criteria can be found in the study of general medicine in University of Ljubljana that sets following criteria for candidates with General Matura (MIZŠ, 2014):

- Success at General Matura – 35 %;
- Final grades in last two years of upper secondary school – 20 %;
- Success at specific subjects on General Matura: Mathematics, Foreign language, one Science subject (Biology, Chemistry or Physics) – 3 x 15 % = 45 %.

Our current admission procedure is governed centrally by office that gathers all the results and implements the selection criteria where needed. Students send their applications in March and they list three study courses they would most like to attend (in the order of preference). At the first step of admission process, only first wish is considered and the list of all study courses with the number of *first wish* candidates is published. If the number of candidates doesn't exceed the number of available places, those candidates will be admitted to the study course after they fulfil other requirements – in most cases this means finishing their upper secondary school (which includes passing the Matura successfully). If there are more candi-

1 Beside subject grades, at the end of each year, students are also given an overall grade that is based on the subject grades and (same as grades) which also goes from 1–5 (1 – Fail, 5 – Excellent).

dates with first wish than places available per study course, selection procedure will be applied.

System aims at transparency and provides fast and straightforward admission for the majority of the population (VPIS UL, 2009–2014). As such it was seen as very satisfactory and didn't have any major revisions. As Table 3 shows, most of the candidates are admitted to the study course listed under their first wish and the proportion is quite stable over time.

Table 3: Proportion of admitted candidates with their first wish (VPIS UL, yearbooks 2009–2014)

Study year	Proportion of admitted (%)
2001/02	79.8
2002/03	81.2
2003/04	77.9
2004/05	81.2
2005/06	79.8
2006/07	78.6
2007/08	79.9
2008/09	83.1
2009/10	82.2
2010/11	80.8
2011/12	80.4
2012/13	78.2
2013/14	76.5

1.9 Importance of first wish

Since choice of study course is a very important decision for each individual, we can assume that the first wish is in fact the study course the candidate would like to study most. We will therefore explore the difference in gender structure between applications and finally admitted candidates with regard to their first wish in study courses where selection criteria had to be applied.

1.10 Existing literature

Search for other recent research on the subject of equity and fairness in transition to higher levels of education in Slovenia produced no results. There is, however, a growing body of literature in Slovenia dealing with gender issues. Mencin Čeplak and Tašner (2009) have established that females are more successful in secondary and tertiary education in Slovenia. While their findings could be compared to our results, we are not interested in comparison of achievement by gender *per se*, but in comparison of school grades compared with national assessment results for same school subjects.

When explaining achievement in languages and reading, Pečjak, Bucik, Peštaj, Podlesek, & Pirc (2010) note that for boys motivational factors are much more important than for girls. This again is interesting for school practice and work on improvements of reading literacy but it is not an issue in fairness and doesn't address the transition to higher level of education.

There's a broad international research literature about external examinations and school grades regarding gender differences. Many authors (Duckworth & Seligman, 2006; Fennema, 1974; Leonard & Jiang, 1995; Maccoby, 1966; Stockard & Wood, 1984; Wentzel, 1988) found greater differences between boys and girls on school grades compared to external examination test results. Of those only Leonard and Jiang (1995) researched predictive validity of Sholastic Aptitude Tests (SAT) scores for high school grade point average (GPA). They found that predictions biased against women but not in a straightforward way that could be corrected by a simple *mathematical fix*. Authors however state different reasons for greater discrepancies between school grades of boys and girls compared to their achievement tests. Van Houtte (2004) for example finds some evidence to support claim that boys' culture is less study oriented which accounts for differences. Tangney, Baumeister, and Boone (2004) point at high self-control as the main factor of better grades which coincides with findings from Duckworth and Seligman (2006), who found that a factor of self-discipline mediated the relationship between gender and achievement. Steinmayr and Spinath (2008) observed that the fact that girls outperform boys can be partly attributed to girls' higher agreeableness and lesser tendency to avoid work. Similar findings were found by Kuhl and Hannover (2012) who found that better school grades girls received could be partly explained with the factor of self-regulated learning.

1.11 Research problem

We will address the question of fairness of selection procedures at both transition points in Slovenian education – from comprehensive nine year schools to upper secondary schools and from upper secondary schools to tertiary education with regard to gender. Since selection procedures at both points are regulated by appropriate regulations and centrally implemented either by ministry of education or

appointed body (Office for admissions in higher education), lack of fairness at transitions can be evidence for sub-optimal governance and research can give insight and evidence for better informed decisions in future.

Since we couldn't find any documents on the rationale for the use of either school grades or external examination results in Slovenian selection procedures, we will assume that selection criteria should be based on achieved knowledge of students and not on other factors such as motivation, working habits, self-discipline, etc. This is clearly an untested assumption, but due to lack of official documentation or research evidence on effects of current selection procedures a necessary one.

1.12 Hypotheses

Since we have two transition points with two selection procedures, we will postulate two similar hypotheses:

- Selection procedure from comprehensive nine year schools to upper secondary schools shows no sign of bias with regard to gender of admitted students.
- Selection procedure from upper secondary schools to tertiary education shows no sign of bias with regard to gender of admitted students.

We are not interested in possible differences between boys and girls but in a differential effect of school grades and achievement scores on selection procedure.

2. Method

2.1 Data and sample

We will use administrative data, gathered through national assessment (NA). NA is a formative assessment and aims at delivering feedback information to student, teacher, school, and system. Through NA database we can obtain both student scores on NA and her/his school grades in last year of school for specific subjects. Since NA is obligatory for all Grade 9 students, our sample will cover whole population that took the test.

2.1.1 National assessment data

NA data are results on standardized tests at the end of lower secondary education (Grade 9) that test student's knowledge according to Slovenian curriculum. All students are tested in Slovene language, mathematics, and the third subject randomly assigned to schools. This third subject for each school is each year selected from a group of four (which are in turn each year selected on national level among all subjects available) and sampling is done in such way to assure that results of sub-

groups are representative for whole Slovenian population. In case of national assessment data in 2013 "third" subjects were: (a) English language, (b) Geography, (c) History, and (d) Engineering & Technology. For research, individual scores on each standardized test for a population of 2013 were used and linked with school grades in the same subjects. School grades range from 1 (fail) to 5 (excellent) with Grade 2 as a first passing grade. Table 4 shows main descriptive statistics for datasets used.

Table 4: Descriptive statistics for national assessment (NA) data

Subject	*N* of students	Mean NA score	*SD* of NA score	*N* of students with school grades
Slovene language	17,217	51.56	16.58	17,141
Mathematics	17,280	55.03	20.02	17,201
English language	4,194	64.05	24.14	4,165
Geography	4,182	63.96	16.86	4,163
History	4,483	47.68	16.90	4,460
Engineering & Technology	4,134	58.35	16.00	4,134

Note. Small proportion of students was tested in German language as their third subject. Since the number of students was small it was not included in analysis.

2.1.2 Data about admission to tertiary education

Data that would allow meaningful insight into transition to tertiary education is not readily available. We will use the data collected by National Examinations Centre in the year 2004 from all existing university study courses at the time that gives insight into an admittance process for the years 1997–2002, of which we will use only the last school year 2001/2002. We will complement this data with data from General Matura for the last few years to show stability in trends and results. Although part of data is relatively old, it does offer population overview over situation and eliminates bias of sampling that other datasets might have. With respect to selection criteria transition to higher education hasn't changed substantively since 1996 and we can assume that most of the findings from 2002 dataset would still apply meaningfully. We will focus our direction to study courses where selection procedures were applied in practice.

Usual analyses, reported by Office for admittance show that selection procedure is neutral in regard to gender structure, since the proportions of finally admitted candidates are not far apart (i.e., for 2013/14 the proportion was 57 % females and 43 % males – VPIS UL, 2014). But if we are interested in effects of selection procedure we must analyze only situations where selection was applied and we should

assume, that first candidate's wish has greater weight than other wishes (which are just alternatives).

We will define the attractiveness of the study course as the ratio of candidates that listed the course in their applications under the first wish and the number of candidates actually admitted. This is important since only the study courses with too many first wish candidates enact the selection criteria and only there can we observe the fairness of the procedure.

2.1.3 Data from General Matura examinations

While it is possible to study on tertiary level if you finished high schools with Vocational Matura, it is also true that study courses where selection criteria is applied will in most cases attract only candidates with General Matura, since they have completed academically more demanding educational track. The last dataset in this research therefore comes from the database on General Matura examinations, where every student at the end of upper secondary school in highest educational track must take General Matura. It is a series of five high stakes tests, where the results are used in admission process to the university. These databases include whole population taking the examinations each year and are accessible through National Examinations Centre. General Matura consists of five external examinations: Slovene language, Mathematics, and first foreign language examinations are obligatory while last two subject examinations are selected by candidate from wide variety of available subjects. For each subject candidate receives grades 1 to 5 (1 – fail, 5 – excellent) or if subject is on higher level 1 to 8. Candidate must pass all five examinations to pass General Matura and his final score is a sum of all five grades. Since up to three examinations can be on higher level, maximum number of points received on General Matura can be 34 (8+8+8+5+5), while minimal passing score for a candidate is 10 (2+2+2+2+2).

2.2 Analyses

Since we will be using large datasets, most differences will be statistically significant on usual levels of significance and not very informative. We will therefore focus on effect sizes as defined by Cohen (1988) and ordinal dominance graphs as defined by Bamber (1975).

Ordinal dominance (OD) graphs offer information which is exact, easily understandable, and allows comparison of ordinal data (Bren & Zupanc, 2010). OD graphs are associated with Mann Whithey's U test and the area under curve can be meaningfully interpreted and compared with other OD graphs (Bamber, 1975). The area under curve (i.e., X) can be literally interpreted in the following way (Jewett, 1983): If we took randomly one subject from each group compared, X would be the

probability that the subject from group A (that's on the x-axis) has higher score or equal to the subject from Group B. We will use this methodology since the variables at hand are clearly ordinal (for school grades we can't assume interval properties) and the methodology in itself is pretty straightforward.

3. Results

3.1 Transition to upper secondary schools – School grades and NA results by gender

Table 5 shows descriptive statistics for school grades and NA results broken down by subject and gender. Effect sizes as defined by Cohen (1988) are calculated to give impression of differences over different metrics. In every situation effect sizes are calculated as

$$(M_{BOYS}-M_{GIRLS})/SD.$$

Therefore, negative values display differences in favor of girls while positive values denote differences in favor of boys.

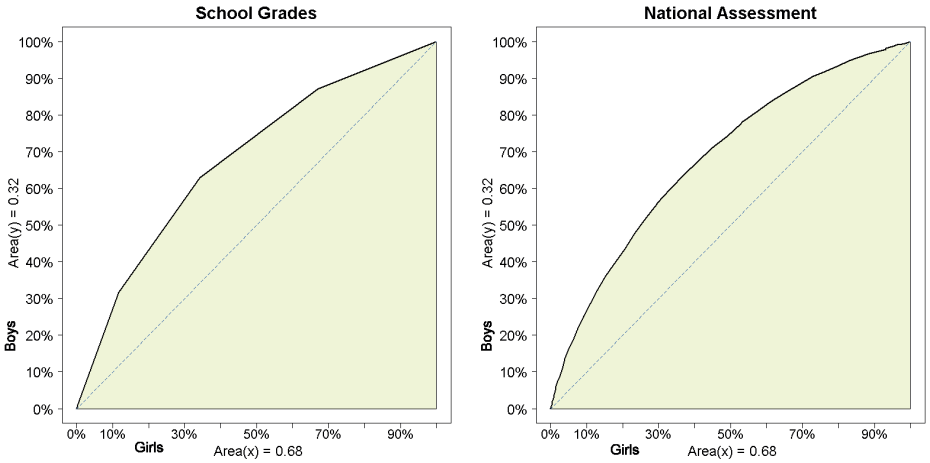
Table 5: Average school grades and national assessment results for 2013 by gender

Subject	Gender	*N* (sch. Grades)	Mean sch. grade	*SD* sch. grade	Effect size	*N* (Nat. assess.)	Mean NA score	*SD* NA score	Effect size
Slovene	F	8,532	3.87	1.01		8,380	57.0	15.5	
Slovene	M	9,006	3.17	1.04	-0.68	8,837	46.4	15.9	-0.68
Math	F	8,528	3.52	1.12		8,400	56.1	20.0	
Math	M	9,006	3.15	1.13	-0.33	8,880	54.1	20.0	-0.10
English	F	2,023	3.81	1.10		1,994	65.5	23.5	
English	M	2,130	3.39	1.10	-0.38	2,103	63.1	24.3	-0.10
Geography	F	1,994	3.92	1.03		1,959	65.4	16.7	
Geography	M	2,166	3.45	1.08	-0.45	2,142	62.9	16.9	-0.15
History	F	2,174	3.77	1.12		2,139	49.2	16.5	
History	M	2,277	3.43	1.15	-0.30	2,219	46.7	17.0	-0.15
Engineering & Technology	F	2,027	4.43	0.77		1,967	57.3	15.1	
Engineering & Technology	M	2,107	4.12	0.93	-0.36	2,044	59.9	16.6	0.16

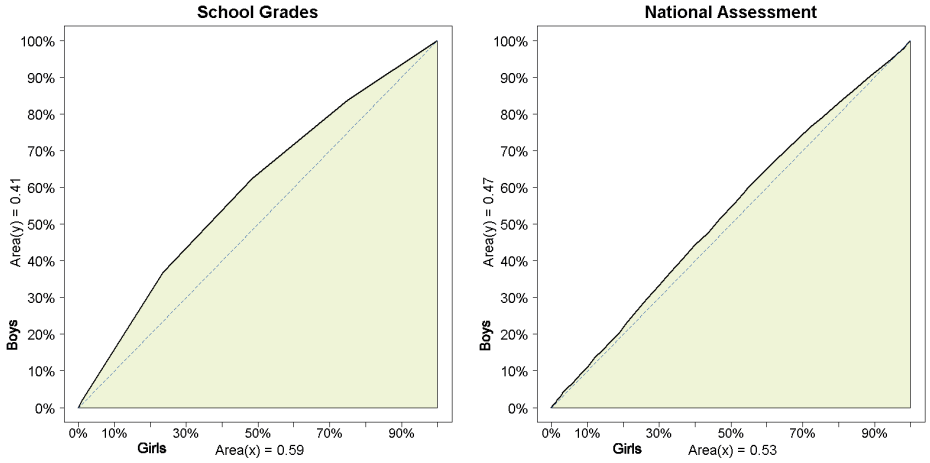
When we compare school grades and NA test results through effect sizes of the differences, we can see that differences coincide in Slovene language (d = -0.68 in both cases) while they differ in all other subjects and typically school grades favor girls much more than NA scores do.

We will make similar comparison graphically with the help of ordinal dominance graphs.

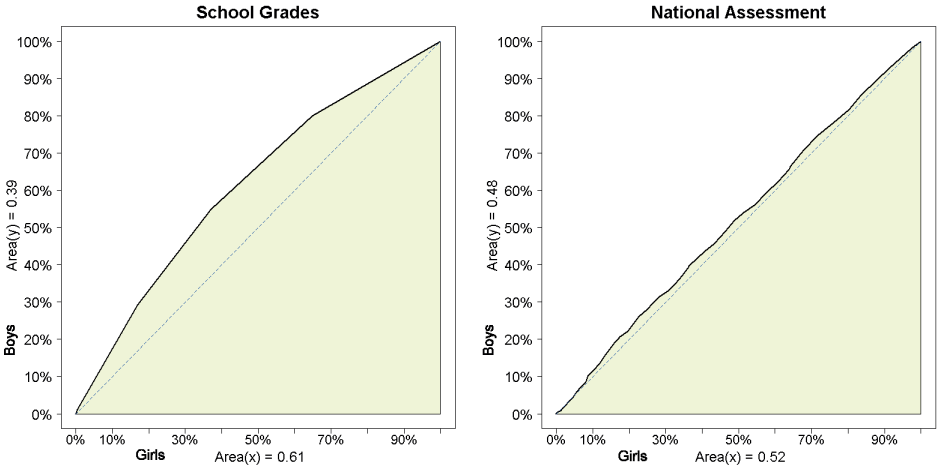
Figures 1 & 2: Ordinal dominance graphs for Slovene language in Grade 9 2013



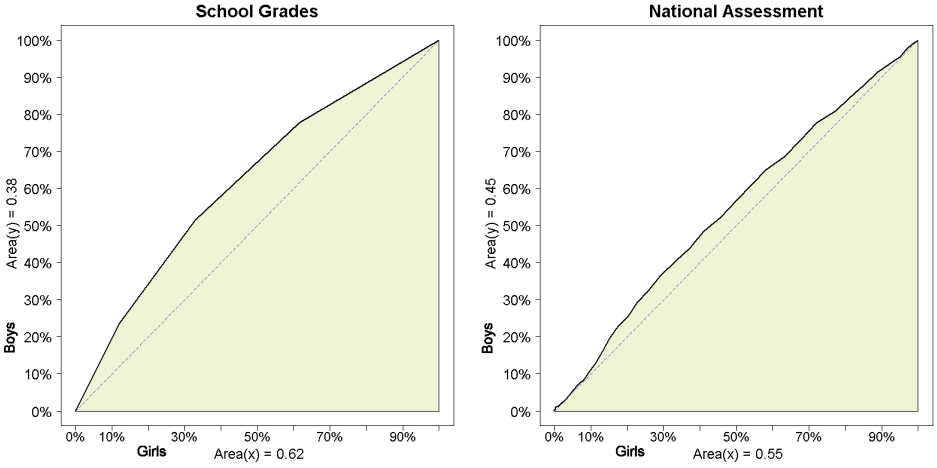
Figures 3 & 4: Ordinal dominance graphs for Mathematics in Grade 9 2013



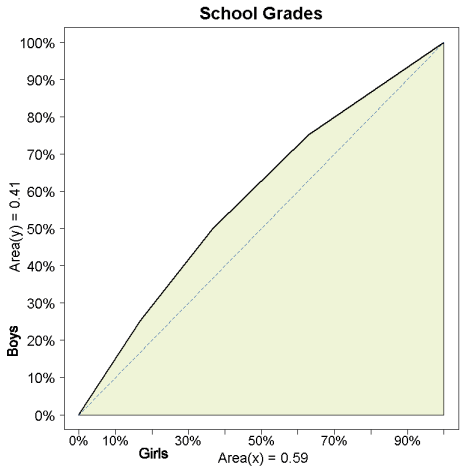
Figures 5 & 6: Ordinal dominance graphs for English language in Grade 9 2013

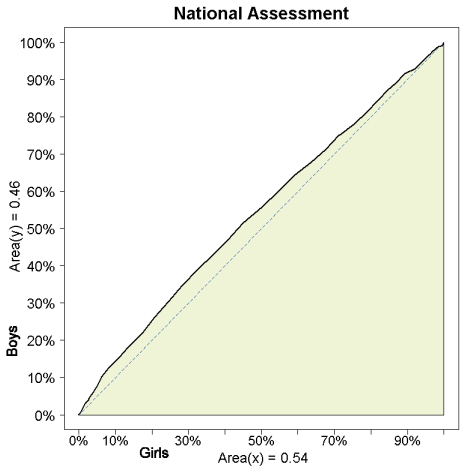


Figures 7 & 8: Ordinal dominance graphs for Geography in Grade 9 2013

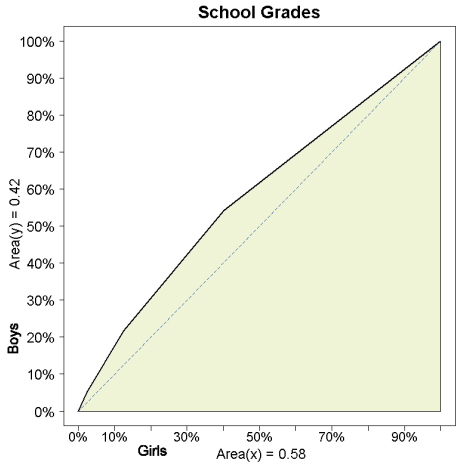


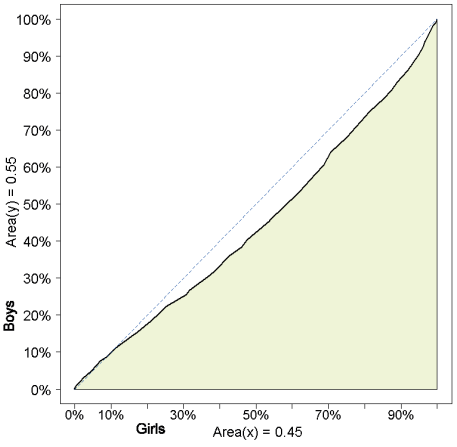
Figures 9 & 10: Ordinal dominance graphs for History in Grade 9 2013





Figures 11 & 12: Ordinal dominance graphs for school subject Engineering & Technology in Grade 9 2013





Figures 1 and 2 show that in Slovene language girls have on average higher scores on NA and have correspondingly higher school grades, too. Both curves show similar dominance of one group over another. Therefore, both measures behave similarly with regard to gender. In Figures 3 to 10 you can see that school grades favor girls compared to the NA results in most school subjects tested – Mathematics, English, Geography, and History. In Engineering & Technology (Figures 11 and 12) girls get better school grades while boys get better NA results. When comparing school grades and NA scores as criterion for selection, boys stand better chances with NA results and girls would on average profit more from school grades. All differences between boys and girls in OD graphs are statistically significant; this is to be expected due to large sample sizes (Nunnally & Bernstein, 1994). We are more interested in statistical significance of percent differences between OD graphs depicting results of NA and school grades. Statistical significance is in this case dependent only on the sample size and size of difference and with sample sizes in excess of 17,000 any difference larger than 1 % is statistically significant with p value less than 0.01! All pairwise differences between grades and NA results in Figures 3 to 12 for each subject are therefore statistically significant. Apart from Slovene language we demonstrated differences in both measures of student's achievement for all other school subjects. Given our hypothesis the results show possibilities of bias in all school subjects except Slovene language.

3.2 Transition to tertiary education

Table 6 shows information for groups of study courses that had more applications from first wish candidates than there were places available. In school year 2001/02, there were 60 study courses where selection procedure was applied. We can further divide them into three groups according to the actual selection criteria applied. The largest group of study courses (38 out of 60) lists same selection procedure: 60 % of points come from success on Matura examination while 40 % comes from final grades in last two years of upper secondary school. Next group of study courses has different proportions of points from Matura and final grades with some of them additionally pondering certain subjects at Matura. There were also study courses that test special talents in addition to first two criteria – those were sorted into third group of programs.

Table 6: Overview of attractiveness of selected study courses, of enlisted and admitted candidates with their first wish by gender

		Number of candidates with first wish			Number of admitted candidates with first wish		
	Number of programs	All	Male	Female	All	Male	Female
Typical proportion 60/40	38	4,464	1,503	2,961	2,306	745	1,561
Various proportions	19	2,815	832	1,983	1,534	435	1,099
Special tests (arts, music, architecture)	3	730	359	371	279	135	144
Total	60	8,009	2,694	5,315	4,119	1,315	2,804

We can see that in the first two groups where is the majority of study courses the number of female candidates exceeds that of male candidates. While this is interesting, it is not an indication of (un)fairness for we should not assume bias in the reasons and factors that led the candidates to listing (or not listing) certain study course under their first wish.

The first group of study courses has largest number of candidates and will therefore allow most valid observations. In Table 7 we calculated the proportions of admitted candidates by gender for this group of study courses. We notice that the proportions are not equal and in fact are a bit larger for females.

Table 7: Comparison of enlisted and admitted candidates for a group of study courses with 60/40 selection criteria

	ADMITTED		NOT ADMITTED		TOTAL	
	Number	Percent	Number	Percent	Number	Percent
Male	745	49.6	758	50.4	1,503	100
Female	1,561	52.7	1,400	47.3	2,961	100
Total	2,306	51.7	2,158	48.3	4,464	100

The differences in admittance by gender in Table 7 are not great and this can be seen as some indication of fairness, although χ^2 is still statistically significant ($\chi^2 = 3.96$, $p = 0.05$). To see if this differences aggravate with increasing attractiveness of the study course we can also select a group of "most attractive" study courses by arbitrarily selecting a threshold and including all study courses, where the number of applications exceeded the number of available places for a factor of 1.5. In the group of courses with selection criteria 60/40, there were 18 such courses and results on admittance are shown in Table 8.

Table 8: Comparison of enlisted and admitted candidates for a group of "most attractive" study courses with 60/40 selection criteria (enlisted/admitted > 1.5)

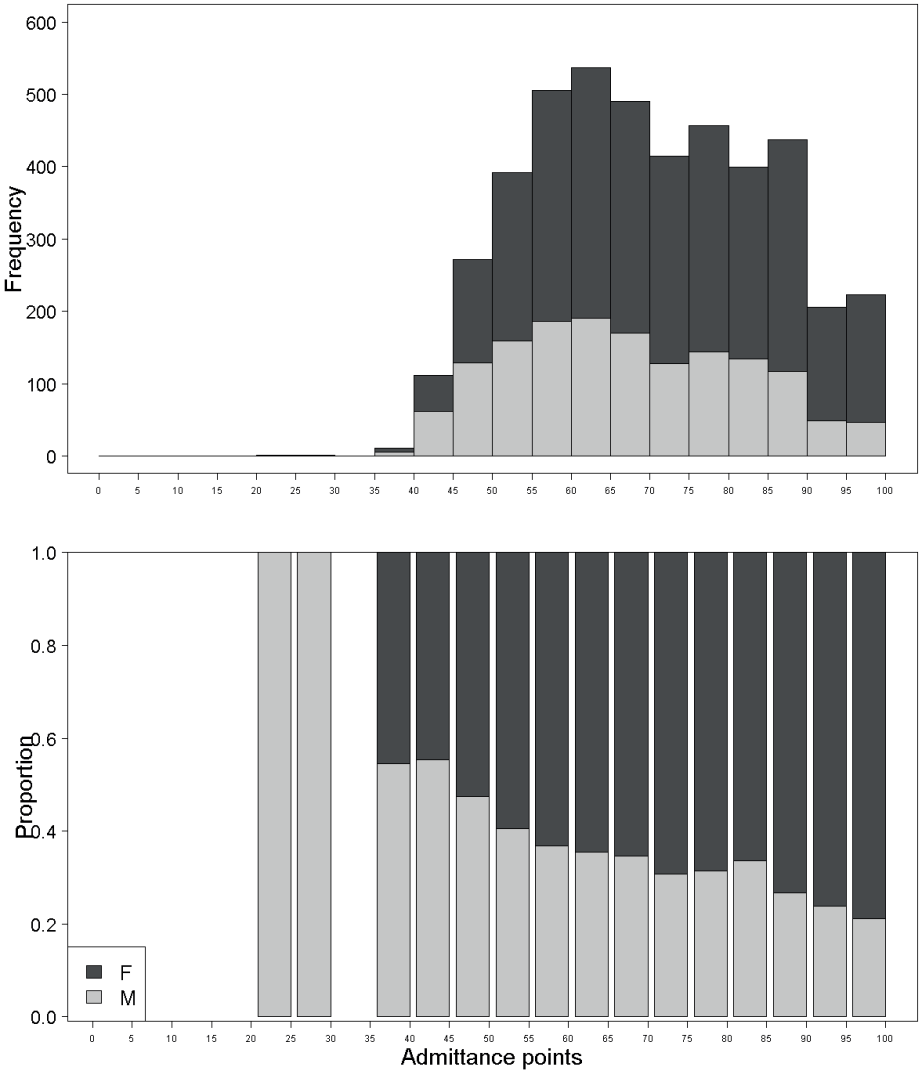
	ADMITTED		NOT ADMITTED		TOTAL	
	Number	Percent	Number	Percent	Number	Percent
Male	174	30.4	399	69.6	573	100
Female	619	39.3	957	60.7	1,576	100
Total	793	36.9	1,356	63.1	2,149	100

In study courses where selection is most strict, there seems to be a bit larger proportion of admitted females compared to males. Differences are statistically significant ($\chi^2 = 14.33$, $p = 0.0002$). The reason for this was simple – female candidates were admitted because they scored more points on selection criteria – success at Matura and final grades from school.

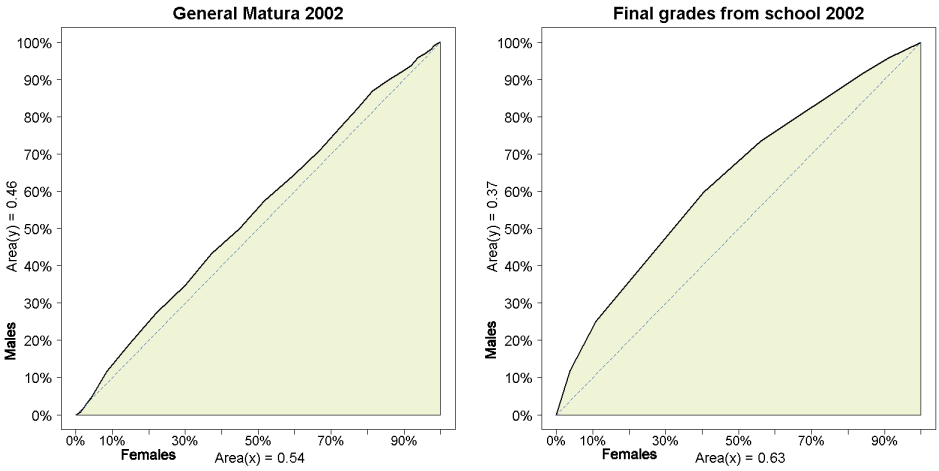
Figure 13 shows the distribution of admittance points for all 4,464 candidates that applied for attractive study courses in group *60/40* from Table 8. You can see the distribution and the relative proportions by gender.

As seen in Figure 13, smaller proportion of males achieved higher number of points compared to females, resulting in a shift of selection criteria in favor of females as demonstrated earlier. We can't deduce from admittance points why this happened, but since 60 % of the points come from success at Matura and 40 % from the sum of final grades in last two years of upper secondary school, we can check both sources of selection criteria separately in the next step. Figures 14 to 17 show ordinal dominance graphs of Matura results and final grades by gender respectively for the admittance results (candidates that in 2002 applied for attractive study courses 60/40) and for all the candidates on General Matura in the last year (2002). Figures 16 and 17 (for General Matura 2013) are added to show stability of differences over time and different cohorts of students.

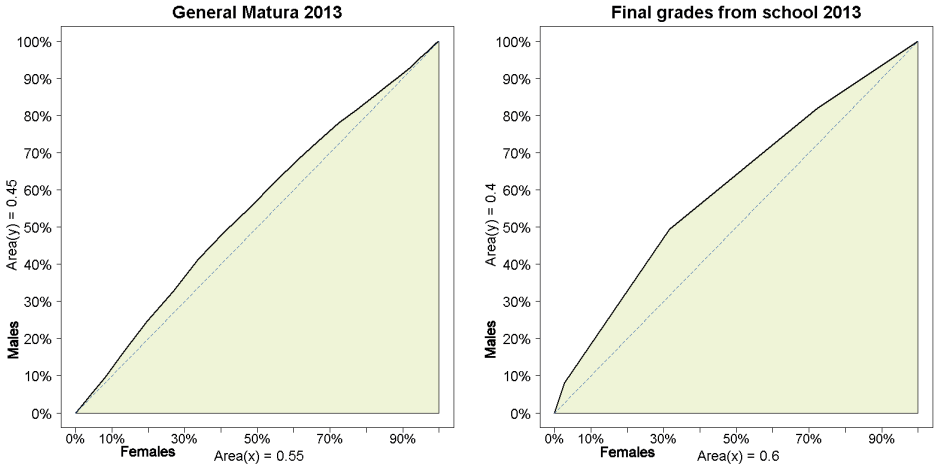
Figure 13: Distribution of admittance points for candidates in group of study courses 60/40 by gender (F = female, M = male)



Figures 14 & 15: Comparison of General Matura results and sum of final grades from upper secondary school by gender for candidates that in 2002 applied for attractive study courses 60/40



Figures 16 & 17: Comparison of General Matura results and sum of final grades from upper secondary school by gender for all candidates in 2013



4. Discussion

4.1 Transition to upper secondary schools – School grades and NA results by gender

It makes sense that school grades don't convey identical information as achievement results. They are obtained in a different manner over longer period of time, are influenced by different factors, etc. ... It is also reasonable that looking on the individual level one could find big differences between someone's student achievement results and her/his associated school grades. However, two large groups of students with similar average achievement should on average have similar school grades in the same school subject. Whatever is the difference between assessment scores and school grades, they should be identical for each subgroup, otherwise there's bias which we should be aware of.

Effect sizes in Table 5 and Figures 1 to 12 show that apart from Slovene, differences in school grades are much larger than differences in NA results. This reveals current problems with transition to upper secondary education in Slovenia. Since 2005, most of the selection criteria rely on school grades and if we assume in line with findings of Zupanc and Bren (2010) that external examinations used in NA are objective and valid measures of student's knowledge, than current system favors girls in the selection. This could to some extent explain relatively large proportion of girls in highest educational track.

Only if we assume that external examination results provide more valid measure of student's achievement, we can further reject the null hypothesis of no bias in selection procedure for admittance to upper secondary schools. What exactly should be corrected in this case remains unclear. We have no documents explaining rationale behind current selection procedure – maybe school grades are better at capturing the essence of criteria intended for selection? Kenney-Benson, Pomerantz, Ryan, and Patrick (2006) point out that girls with similar achievement have higher grades because of their learning strategies. Whether the reason behind differences in school grades are indeed learning strategies or maybe self-discipline (Duckworth & Seligman, 2005), delay of gratification (Silverman, 2003), higher agreeableness, and lesser tendency to avoid work (Steinmayr & Spinath, 2008), gender-specific culture (Van Houtte, 2004), or even gender identity (Vantieghem, Vermeersch, & Van Houtte, 2014), we cannot assess fairness of procedure if rationale for its implementation is not explicitly stated. Without this, we cannot decide which (school grades or achievement scores) show undesirable characteristics. In the absence of a clear rationale for selection criteria, these results offer a starting point to build one in the future or tailor the selection procedure accordingly.

4.2 Transition to tertiary education

While we can see that females outperform males on results from General Matura and on results from final grades in Figures 14 to 17, it is also obvious that most of their advantage in admittance points comes from school grades. If we assume that external examinations are more objective and neutral measure of person's knowledge than school grades and that achievement on external examinations is less "burdened" with other factors like classroom discipline, working habits, motivation, etc., then school grades display bias in favor of females compared with results on General Matura. This is consistent with research literature. Duckworth and Seligman (2006) report in their research that girls consistently received higher school grades while they just marginally outperformed boys on achievement tests. They suggested that school grades were higher because girls were much better than boys in self-discipline at school, which resulted in higher grades. This is consistent with findings of Leonard and Jiang (1995) who report that Sholastic Aptitude Tests (SATs) are biased against women. They came to this conclusion since the predicted GPA in high schools based on SAT scores is for females typically lower than actual GPA.

Differences in performance on achievement tests and school grades are stable and consistent from year to year. Comparison of same data on whole population of last year's candidates on General Matura shows remarkably similar picture to the candidates in admittance procedure of 2002 shown earlier. With sample sizes 4,000 (over 7,000 in case of 2013 data) 1 % change in area under curve is statistically significant with p value under 0.01. Since both cases show much larger differences (9 % & 5 % respectively), we can reject the null hypothesis of no bias of selection criteria by gender.

At this point we could ask ourselves if additional factors like self-discipline, learning strategies, delay of gratification, lesser tendency to avoid work, or gender identity, presumably measured in school grades besides achievement, should be used in selection criteria or not. Under assumption that factors outside achievement shouldn't influence selection, we can demonstrate unfairness of the current selection procedures. But an opposite view could also be adopted and since one of the measures displays much larger differences by gender than the other, you can always, depending on the view adopted, question the fairness of one of the measures.

This is in fact a question of validity and this research should serve to appropriate authorities in Slovenia to review the rationale behind the selection and decide if selection procedures should be improved. In case of transition to tertiary level of education for example Geiser and Santelices (2007) point out that high school grades display excellent predictive validity for student success, even greater than achievement tests.

If we assume that selection should be based on achievement only, we can detect unfairness of the transition to tertiary educations as current procedure allows easier transition for females in comparison to their male peers. Since study cours-

es didn't declare their incentives behind their selection criteria, this assumption for now remains untestable.

4.3 Generalizability of results

Although inferences about the transition to tertiary education are very interesting, they do come from quite an old data and further research on new data should be repeated to see current trends. However, both selection procedures and both sources of data used in selection (Matura results and final grades) are pretty stable in time as can be seen in Table 9:

Table 9: Means of General Matura results and of final grades in last year of upper secondary school for years 2007–2013

	Number		Mean of Success at Gen. Matura		Mean of final grade in last year of school	
	Females	Males	Females	Males	Females	Males
2006/07	4,660	3,009	19.94	19.20	3.82	3.47
2007/08	4,476	2,892	20.16	19.45	3.83	3.49
2008/09	4,472	2,804	19.90	19.33	3.85	3.50
2009/10	4,221	2,693	19.66	19.16	3.90	3.54
2010/11	3,993	2,598	19.98	19.30	3.87	2.58
2011/12	3,763	2,500	20.50	19.79	3.89	3.54
2012/13	3,924	2,411	20.35	19.55	3.93	3.60

Differences by gender seem pretty stable from year to year which makes results from an older dataset more generalizable, although for greater insight new data should be collected and compared to these.

5. Conclusion

Both points of transition to higher level of education in Slovenia demonstrate some level of bias. This article avoided prevalent focus on statistical testing of the differences, since the datasets included whole populations and in very large samples most of the observed differences are statistically significant anyway. There was also less focus on differences between boys and girls, as this paper tried to show the differences between school grades and achievement scores, and how they translate to fairness in selection. This is especially true for governance of transition to upper secondary schools where reliance on school grades as main selection criterion creates heavy bias for one group over another. There is also an unanswered question

of validity of each selection criterion for high stakes decision that should be explored and answered before the criterion is used for selection any further.

Whether the results presented here provide evidence of unfair selection and sub-optimal governance remains to be decided by further research. The lack of appropriate and up-to-date data was a problem of this research and is most probably a reason for non-existent body of research on Slovenian selection procedures in the first place. To evaluate governance, there should also be documents providing rationale and supporting evidence for the decisions made. In case of selection criteria on each point of transition in Slovenian education, those documents are missing and decisions are not evidence based. Research provided insights into the problem and gave evidence to authorities to make informed decisions and improve governance.

References

Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, *12*(4), 387–415.

Bren, M., & Zupanc, D. (2010). Ordinal statistics to answer ordinal questions. In A. Graovac, B. Pokrić, & V. Smrečki (Eds.), *The 25th international course & conference MATH/CHEM/COMP 2010, program & book of abstracts* (p. 11). Zagreb, Croatia: Ruđer Bošković Institute.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Coward, R. (2010). Educational governance in the NHS: A literature review. *International Journal of Health Care Quality Assurance*, *23*(8), 708–17. doi:http://dx.doi.org/10.1108/09526861011081840

De Coster, I., Forsthuber, B., Oberheidt, S., Parveva, T., & Glass, A. (2008). *Higher education governance in Europe: Policies, structures, funding and academic staff*. Retrieved from http://eric.ed.gov/?id=ED540821

Duckworth, A. L., & Seligman, M. E. (2005). Self-Discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, *16*(12), 939–944.

Duckworth, A. L., & Seligman, M. E. (2006). Self-Discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology*, *98*(1), 198–208. doi:10.1037/0022–0663.98.1.198

Eurydice. (2014). *Structure of the national education system 2012/13*. Retrieved from https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/Slovenia:Overview

Fazekas, M., & Burns, T. (2012). *Exploring the complex interaction between governance and knowledge in education* (OECD Education Working Papers No. 67). Retrieved from http://www.oecd–ilibrary.org/education/exploring–the–complex–interaction–between–governance–and–knowledge–in–education_5k9flcx2l340–en

Fennema, E. (1974). Mathematics learning and the sexes: A review. *Journal for Research in Mathematics Education*, *5*(3), 126–139.

Gaber, S., Cankar, G., Marjanovič Umek, L., & Tašner, V. (2012). The danger of inadequate conceptualisation in PISA for education policy. *Compare: A Journal of Comparative and International Education, 42*(4), 647–663.

Geiser, S., & Santelices, M. V. (2007). *Validity of high-school grades in predicting student success beyond the freshman year: High-School record vs. standardized*

tests as indicators of four-year college outcomes. (Research & Occasional Paper Series: CSHE.6.07). Retrieved from University of California, Center for Studies in Higher Education website: http://www.cshe.berkeley.edu/publications/validity-high-school-grades-predicting-student-success-beyond-freshman-yearhigh-school

Hénard, F., & Mitterle, A. (2010). *Governance and quality guidelines in higher education. A review of governance arrangements and quality assurance guidelines*. Paris, France: OECD.

Jewett, J. P. (1983). *A comparative study of the statistical properties of the OD curve and the Q-Q plot*. Vermillion, SD: University of South Dakota.

Kenney-Benson, G. A., Pomerantz, E. M., Ryan, A. M., & Patrick, H. (2006). Sex differences in math performance: The role of children's approach to schoolwork. *Developmental Psychology*, *42*(1), 11–26. doi:10.1037/0012–1649.42.1.11

Krucken, G. (2011). *A European perspective on new modes of University Governance and Actorhood*. (Research & Occasional Paper Series: CSHE.17.11). Retrieved from University of California, Center for Studies in Higher Education website: http://www.cshe.berkeley.edu/european-perspective-new-modes-university-governance-and-actorhood

Kuhl, P., & Hannover, B. (2012). Differenzielle Benotungen von Mädchen und Jungen: Der Einfluss der von der Lehrkraft eingeschätzten Kompetenz zum selbstgesteuerten Lernen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *44*(3), 153–162. doi:10.1026/0049–8637/a000066

Leonard, D. K., & Jiang, J. (1995). *Gender bias in the college predictions of the SAT*. Retrieved from http://eric.ed.gov/?id=ED383736

Maccoby, E. E. (1966). *The development of sex differences*. Stanford, CA: Stanford University Press.

Mencin Čeplak, M., & Tašner, V. (2009). Spolne neenakosti v izobraževanju [Gender inequalities in education]. In V. Tašner (Ed.), *Brez spopada: kultur, spolov, generacij* [Without encounter: Cultures, gender, generations] (pp. 103–116). Ljubljana, Slovenia: Pedagoška fakulteta, Univerza v Ljubljani.

MIZŠ – Ministry for Education, Science and Sport. (2014). *Razpis za vpis na dodiplomske in enovite magistrske* študijske *programe v* študijskem *letu 2014/2015 – Univerza v Ljubljani, Univerza v Mariboru, Univerza na Primorskem, Univerza v Novi Gorici, javni in koncesionirani samostojni visokošolski zavodi*. Ljubljana, Slovenia: MŠŠ. Retrieved from http://www.mizs.gov.si/si/delovna_podrocja/direktorat_za_visoko_solstvo/sektor_za_visoko_solstvo/razpis_za_vpis/razpis_za_vpis_javni_in_koncesionirani/#c20128]

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.

OECD – Organisation for Economic Co-operation and Development. (2003). *Education policy analysis 2003*. Washington, DC: Organization for Economic Cooperation & Development.

OECD – Organisation for Economic Co-operation and Development. (2007). *PISA 2006 competencies for tomorrow's world. Volume 1: Analysis*. Paris, France: OECD Publishing.

Pečjak, S., Bucik, N., Peštaj, M., Podlesek, A., & Pirc T. (2010). Bralna pismenost ob koncu osnovne šole – ali fantje berejo drugače kot dekleta? *Sodobna pedagogika*, *61*(1), 86–102.

Silverman, I. W. (2003). Gender differences in delay of gratification: A meta-analysis. *Sex Roles, 49*(9–10), 451–463.

Slavin, R. E. (2002). Evidence-Based education policies: Transforming educational practice and research. *Educational Researcher*, *31*(7), 15–21. doi: 10.3102/0013189X031007015

Steinmayr, R., & Spinath, B. (2008). Sex differences in school achievement: What are the roles of personality and achievement motivation? *European Journal of Personality*, *22*(3), 185–209. doi: 10.1002/per.676

Stockard, J., & Wood, W. (1984). The myth of female underachievement: A reexamination of sex differences in academic underachievement. *American Educational Research Journal*, *21*(4), 825–838.

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of personality*, *72*(2), 271–324.

Van Houtte, M. (2004). Why boys achieve less at school than girls: The difference between boys' and girls' academic culture. *Educational Studies*, *30*(2), 159–173.

Vantieghem, W., Vermeersch, H., & Van Houtte, M. (2014). Transcending the gender dichotomy in educational gender gap research: The association between gender identity and academic self-efficacy. *Contemporary Educational Psychology*, *39*(4), 369–378.

VPIS UL – Vpisna prijavno informacijska služba Univerze v Ljubljani (2009). *Analiza prijave in vpisa,* študijsko *leto 2008/2009* [Applications and admissions analyses, study year 2008/2009]. Ljubljana, Slovenia: VPIS UL.

VPIS UL – Vpisna prijavno informacijska služba Univerze v Ljubljani (2010). *Analiza prijave in vpisa,* študijsko *leto 2009/2010* [Applications and admissions analyses, study year 2009/2010]. Ljubljana, Slovenia: VPIS UL.

VPIS UL – Vpisna prijavno informacijska služba Univerze v Ljubljani (2011). *Analiza prijave in vpisa,* študijsko *leto 2010/2011* [Applications and admissions analyses, study year 2010/2011]. Ljubljana, Slovenia: VPIS UL.

VPIS UL – Vpisna prijavno informacijska služba Univerze v Ljubljani (2012). *Analiza prijave in vpisa,* študijsko *leto 2011/2012* [Applications and admissions analyses, study year 2011/2012]. Ljubljana, Slovenia: VPIS UL.

VPIS UL – Vpisna prijavno informacijska služba Univerze v Ljubljani (2013). *Analiza prijave in vpisa,* študijsko *leto 2012/2013* [Applications and admissions analyses, study year 2012/2013]. Ljubljana, Slovenia: VPIS UL.

VPIS UL – Vpisna prijavno informacijska služba Univerze v Ljubljani (2014). *Analiza prijave in vpisa,* študijsko *leto 2013/2014* [Applications and admissions analyses, study year 2013/2014]. Ljubljana, Slovenia*: VPIS UL.*

Wentzel, K. R. (1988). Gender differences in math and english achievement: A longitudinal study. *Sex Roles*, *18*(11–12), 691–699. doi: 10.1007/BF00288054

Willingham, W. W., & Cole, N. S. (1997). *Gender and fair assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Zupanc, D., & Bren, M. (2010). Inflacija pri internem ocenjevanju v Sloveniji [Grade inflation in Slovenia]. *Sodobna pedagogogika, 61*(3), 208–228.