

Sirkku Kupiainen, Jukka Marjanen & Jarkko Hautamäki

The problem posed by exam choice on the comparability of results in the Finnish matriculation examination

Abstract

The article by Kupiainen, Marjanen and Hautamäki focuses on the upper secondary matriculation examination in Finland as a school leaving and university entrance examination. The presented research addresses the question of whether increased choice of the subject-specific examinations has the potential to undermine the comparability of examination results and to direct students' choices not only in the examination but already beforehand at school. The authors refer to Finland's tradition of more than 160 years of a national examination connecting the academic track of upper secondary schools with universities. The authors explain the Finnish system by describing the adoption of a course-based (vs. class- or year-based) curriculum for the three-year upper secondary education and the subsequent reforms in the matriculation examination. This increases students' choices considerably with regard to the subject-specific exams included in the examination (a minimum of four). As a result, high-achieving students compete against each other in the more demanding subjects while the less able share the same normal distribution of grades in the less demanding subjects. As a consequence, students tend to strategic exam-planning, which in turn affects their study choices at school, often to the detriment of the more demanding subjects and, subsequently, of students' career opportunities, endangering the traditional national objective of an all-round pre-academic upper secondary education.

This contribution provides an overview of Finnish upper secondary education and of the matriculation examination (cf. Klein, 2013) while studying three separate but related issues by using data from several years of Finnish matriculation results:

- *the relation of the matriculation examination and the curriculum;*
- *the problems of comparability vis-à-vis university entry due to the increased choice within the examination;*

Sirkku Kupiainen M.Sc. (corresponding author) · Jukka Marjanen M.Ed. · Prof. Dr. Jarkko Hautamäki, Centre for Educational Assessment, Department of Teacher Education, University of Helsinki, PL 9 (Siltavuorenpenger 3A), 00014 University of Helsinki, Finland

e-mail: sirkku.kupiainen@helsinki.fi
jukka.marjanen@helsinki.fi
jarkko.hautamaki@helsinki.fi

- *the relations between students' examination choices and their course selection and achievement during upper secondary school.*

Keywords

Exit exams; Grade comparability; Impact of choice on comparability; Teachers' vs. national grades

Das Problem der Wahlmöglichkeiten von Prüfungsfächern hinsichtlich der Vergleichbarkeit von Ergebnissen der zentralen Abschlussprüfung in Finnland

Zusammenfassung

Der Artikel von Kupiainen, Marjanen und Hautamäki konzentriert sich auf die zentrale Abschlussprüfung der Sekundarstufe II in Finnland als eine Schulabschluss- und Hochschulzugangsprüfung. Die Studie geht der Frage nach, ob die gestiegenen Auswahlmöglichkeiten der fachspezifischen Prüfungen die Vergleichbarkeit der Prüfungsergebnisse und die Wahl der Schülerinnen und Schüler nicht nur in der Prüfung, sondern bereits während der Schulzeit beeinflussen kann. Es wird Bezug auf Finnlands mehr als 160 Jahre lange Tradition zentraler Abschlussprüfungen am Übergang zwischen Sekundarstufe II und Hochschulzugang genommen. Die Autorengruppe erläutert das finnische System hinsichtlich der Einführung eines kursbasierten (vs. klassen- oder jahrgangsstufenbasierten) Curriculums für die dreijährige Sekundarstufe II und bezüglich der anschließenden Reformen der zentralen Abschlussprüfung, durch welche die Auswahlmöglichkeiten von Schülerinnen und Schüler für die fachspezifischen Prüfungen (mindestens vier) wesentlich erhöht wurden. Die Ergebnisse zeigen, dass leistungsstarke Schülerinnen und Schüler in anspruchsvolleren Fächern miteinander konkurrieren, während leistungsschwächere Schülerinnen und Schüler dieselben Noten in weniger anspruchsvollen Fächern erlangen. Als Konsequenz daraus neigen Schülerinnen und Schüler zu einer strategischen Prüfungsplanung, welche wiederum – oft zum Nachteil der anspruchsvolleren Fächer und der anschließenden Karrierechancen – Auswirkungen auf ihre Fächerwahl in der Schule hat, was die traditionellen, nationalen Ziele einer ganzheitlichen vorakademischen Sekundarschulbildung gefährdet.

Dieser Beitrag bietet einen Überblick über das höhere Sekundarschulwesen in Finnland und die zentrale Abschlussprüfung (vgl. Klein, 2013). Unter Verwendung von Daten finnischer Abschlussprüfungen aus mehreren Jahren werden drei separate, jedoch miteinander verknüpfte Aspekte untersucht:

- *der Zusammenhang zwischen Abschlussprüfung und Curriculum;*
- *die Problematik der Vergleichbarkeit hinsichtlich des Hochschulzugangs aufgrund der vergrößerten Auswahlmöglichkeiten bei der Prüfung;*

- *der Zusammenhang zwischen der Prüfungswahl der Schülerinnen und Schüler und ihrer Kurs-Auswahl mit der Leistung während der Sekundarstufe II.*

Schlagworte

Abschlussprüfungen; Notenvergleichbarkeit; Einfluss der Kursauswahl auf Vergleichbarkeit; Benotung der Lehrkräfte vs. Ergebnisse zentraler Prüfungen

1. Background

Upper secondary exit exams bridge the gap between school and university, marking successful passing of the first and acting as a gate-keeper for the second (Noah & Eckstein, 1992). This double role of the exam is especially salient in countries where the share of the age cohort passing the exam exceeds that of students accepted to higher education. When exam results play a prominent role in student admission, the high stakes of the examination are acute and set specific requirements to the comparability of the results of the exams in the different subjects and across years (cf. Béguin, 2000; Coe, 2008; Coe, Searle, Barmby, Jones, & Higgins, 2008).

Exit exams are often seen to improve student performance (Bishop, 1998; Bishop, Mañe, & Bishop, 2001; Jürges, Schneider, & Büchel, 2003; Jürges, Schneider, Senkbeil, & Carstensen, 2012), apparently through a process where students transfer the extrinsic requirements of the school and the society into personal intrinsic motivation (cf. Ryan & Deci, 2000). Pedagogically oriented literature, on the other hand, stresses the adverse effects exams can have on teachers and on instruction (Klein & van Ackeren, 2011), on intrinsic motivation (Jürges et al., 2012), and on students' achievement under the pressure of testing (Amrein & Berliner, 2002; Reardon, Atteberry, Arshan, & Kurlander, 2009).

The introduction of statewide exit exams in many of the German federal states after the 'PISA shock' of the early 2000s (Waldow, 2009) has brought up a new surge of research on exit exams in Germany. The shift from the earlier school-based *Abitur* to the new form of statewide exit exams, even if still differing widely in scope and degree of standardization, have prompted a wide spectrum of research from localized studies comparing the time before and after the change (e.g., Maag Merki, 2011) or focusing on the validity of individual exit exams vis-à-vis tests deriving from other sources (e.g., Kahnert, Eickelmann, Bos, & Endberg, 2012) to looking for differences in the type and level of standardization of the exit exams of countries with a longer tradition of centralized exams from a governance and accountability point of view (e.g., Klein & van Ackeren, 2011).

The highly centralized Finnish matriculation examination with its tightly subject-bound exams (see next Section) is one of the exam systems examined by Klein and van Ackeren (2011) in their article and even more broadly by Klein (2013) in her dissertation. Yet, the topical questions facing the examination are not so much

those of the current German discussion but those of the current British research regarding the lacking comparability of the different examinations of the General Certificate of Secondary Education (GCSE) (Coe, 2008; Coe, et al., 2008) and the Hungarian Matura (Csapó, 2009). Like for the GCSE and the Hungarian Matura, the most acute problem regarding the Finnish matriculation examination is the examination comprising a great number of subject-specific exams, each taken by a skewed share of the examinees, while the results are treated as comparable across the exams and the body of examinees when students apply for tertiary education. The situation is relatively new, due to recent reforms in both upper secondary studies and in the matriculation examination, to be elaborated further in Section 2.

The Finnish matriculation examination complies well with John Bishop's (1998) construct *Curriculum-Based External Exit Exam Systems* (CBEEES) which he sees to be the most germane for advancing student achievement:

- Student's accomplishment in the exam has real consequences.
- The exam defines achievement relative to an external standard.
- The exams are organized by discipline and keyed to the content of specific course sequences.
- The exams signal multiple levels of achievement in the subject.
- The exams cover almost all secondary school students.

Accordingly, we have chosen Bishop's CBEEES as a frame of reference for this first explorative study of the impact of the recent reforms of the structure and syllabus of the Finnish upper secondary schools and of the matriculation examination on students' course and exam choices in the matriculation examination, and on the comparability of the exam results.

2. Upper secondary education in Finland

The matriculation examination, a cherished 160-year-old tradition (Kaarninen & Kaarninen, 2002; Vuorio-Lehti, 2006), is the only high-stakes test in the Finnish education system. The examination marks the exit from the general or academic (vs. vocational) track of upper secondary education and provides a formal qualification for entering university. The grades attained in the examination are taken into account when selecting students to the different faculties, in most cases on side of a separate entry examination. The adoption of the nine year comprehensive school in the 1970s increased the share of students entering the general track schools from about 30 % to closer to 60 %, leading to an increased heterogeneity of students aiming at matriculation and tertiary education. To accommodate to this growing heterogeneity, increased choice was adopted both in the curriculum and in the matriculation examination.

2.1 General upper secondary studies

During the 1990s and early 2000, the traditional three-year structure of the general upper secondary curriculum was abandoned and the syllabus of each subject was divided into independent six-week-long courses, offered successively within five periods during the school year, ending with course-specific exams at the end of the period. Students are to build their own syllabus from these courses for the three to four years of studies, most often with five to six courses per period. Students' choice of courses is governed by the distribution of lesson hours stated in a Government decree (FNBE, 2003, pp. 252–258) with the compulsory syllabus comprising 18 subjects¹. On side of these, many students study additional foreign languages at either the advanced (continuing from the comprehensive school) or at the basic level. Most schools also offer courses in a variety of subjects from Information and communication technologies to Drama.

The requirement for graduation is 75 courses. Of these, 47 are mandatory (51 for students of A-level math). The number of mandatory courses per subject varies from 1 (e.g., chemistry, philosophy, physics, psychology, and health education) to 10 (A-level mathematics), leading to widely varying personal syllabi. In addition to these, the matriculation exam in each subject is based on 'national specialization courses' which all schools have to offer regularly to not impede students' examination plans (FNBE, 2003, p. 255). The number of these courses also varies according to subject. In total, the number of courses covered by the exams in the matriculation examination varies from 3 (health education) to 13 (A-level mathematics).

2.2 The matriculation examination

The Finnish matriculation examination largely conforms to Bishop's (1998) definition for CBEEES (see Section 1): (a) The examination results provide additional credit to university entrance examinations, (b) the separate exams are centrally compiled, administered, and marked (after an initial marking by the teacher according to a centrally prepared guidelines), (c) there is a separate exam for each academic subject and study level (e.g., A- and B-level math, A-, B-, and C-level languages), covering a pre-ordained number (3–13) of courses, (d) a seven point Gaussian scale is superimposed for each exam after an initial criterion-based as-

1 Finnish/Swedish/Sami as mother tongue and/or language of instruction, Swedish/Finnish as the second national language (A- or B-level, beginning at Grade 3 or 7), English as the compulsory foreign language beginning at Grade 3 or 5, Mathematics, Biology, Geography, Physics, Chemistry, History, Social studies, Philosophy, Psychology, Religion or Ethics, Health education, Physical education, Music, Arts, and Educational and vocational guidance.

assessment², and (e) the exam is taken by all general track students (the matriculation examination acts as the official certificate of upper secondary studies together with a final report card compiling grades for all courses). In addition, the results provide feedback to the schools regarding the implementation of the curriculum (Lukiolaki, 1998; Klein & van Ackeren, 2011).

To accommodate the growing heterogeneity of the examinee body, three major reforms have been implemented since the mid-1990s; each weakening the comparability of the results of the different exams due to splitting the examinees into subgroups of differing ability level: In 1994, students were offered the possibility to split their examination to three separate occasions (e.g., spring/fall/spring); in 2005, the mandatory status of the second national language exam (Swedish/Finnish) was waived; and in 2006, the earlier integrated exam for all natural and socio-humanistic science subjects was divided into separate exams for each of the 10³ subjects.

As a result, the matriculation examination is today a compilation of 39 exams in 25 subjects⁴, organized two times a year (spring/fall). Only the exam in Finnish/Swedish (mother tongue or the language of education) is compulsory. In addition, three other compulsory exams are to be chosen from among (a) mathematics, (b) foreign language, (c) the second national language, and (d) science and humanities subjects. One of the exams has to present advanced (A) level. In addition to these, the examinee is free to choose as many optional exams as s/he wishes. The exams are arranged during a three week period in March/September with a six-hour session for each exam, mainly calling for essay-type answers. The majority of exams are taken in the spring of the third (last) year of school, even if today only 5 % of students sit for all exams in the spring of the last year of school, the way everyone did twenty years ago.

Passing the matriculation examination marks eligibility for tertiary education. Yet, only students with the best grades in mathematics and science will be accepted directly into the fields of science, technology, engineering, and mathematics (STEM) while all other selection is through entrance examinations (EF, 2012). As part of the credits awarded on the basis of the matriculation examination are often not tied to specific exams, students may profit from trying to maximize their grades by choosing exams in which their chance of getting a good grade is more probable. Hence, for students not aiming into the STEM fields, focusing already at school on the socio-humanistic subjects with relatively few courses might be a rational strategy.

2 The share of each grade differs slightly by year and by subject but the robust guideline divides the grades from failed (I) to excellent (L) as: Improbatur (I) 5 %, Approbatur (A) 11 %, Lubenter (B) 20 %, Cum Laude Approbatur (C) 24 %, Magna Cum Laude Approbatur (M) 20 %, Eximia (E) 15 %, and Laudatur (L) 5 % (FMEB, 2015a).

3 The actual number of exams is 12 as there are alternative exams for Lutheran and Orthodox religion, and Ethics, but the students can only choose the one they have studied at school.

4 The higher number of exams than subjects reflects the different levels of syllabi and, consequently, of exams in mathematics, languages, and religion/ethics.

3. The present study

Despite the scope of the reforms described above and their centrality in selecting students for tertiary education, no earlier research on the possible effects of the reforms on the comparability of the exam results has been executed. However, the high mean age of Finnish students entering university, largely due to the slow transfer from upper secondary to tertiary education, led the Finnish Government of 2011–2015 to state in its action plan that “the matriculation examination will be developed ... to allow for [its] wider utilization in selecting students for tertiary education by improving the comparability of the exam grades” (VNK, 2011, p. 33).

Reflecting this goal, the present study set to explore the comparability of the matriculation examination grades awarded in the different subjects, changes over time in students’ exam choices, and the relation of students’ exam choice and success in the matriculation examination to their course choices and success in the upper secondary school.

3.1 Research questions

Three research questions were set:

Research Question 1: Do students sitting for different exams differ in their overall attainment in such a way that the same grade (from *improbatur* to *laudatur* according to the distribution presented in Footnote 2) awarded in the different exams represents different levels of overall competence, endangering the comparability of the grades when used as indicators for general academic proficiency at student selection for tertiary education?

Hypotheses 1: It is expected that as the Gaussian distribution of grades takes into account only the attainment of the students sitting for that specific exam⁵, students sitting for exams in the more demanding subjects with a greater number of courses (e.g., A-level mathematics and Physics) get in these exams lower grades than they get in the exams that they share with a more heterogeneous examinee body. This also regards A-level languages other than English but as the number of students sitting for these language exams is more limited as they are not offered in all municipalities, we will center here on the STEM subjects.

Research Question 2: Are there systematic changes in time (2006–2012) in students’ exam choices, indicating that the relative difficulty of the different exams or – more accurately – of the probability to receive a better grade in an exam taken by less able examinees, guides students’ choices?

Hypotheses 2: For the same reason as for Hypotheses 1, it is expected that weaker students look for subjects and, consequently, exams in which their probability of getting higher-than-expected grades are bigger by avoiding those favored

5 It has to be noted that after the initial results of the study were shown to the Finnish Matriculation Examination Board, steps have been taken to begin to alleviate this problem (see Marjanen, 2015).

by the best students. This is expected to have an impact on the popularity of subjects and exams, creating a trend where ever better students become the weakest, trying in their turn to avoid getting the lowest grades by choosing subjects and exams with the lowest number of courses.

Research Question 3: Are there systematic differences between schools in the relation of students' upper secondary course grades and their grades in the respective matriculation exams?

Hypotheses 3: As there is no systematic data on the course-level choices and grades in the Finnish upper secondary schools, no clear-cut hypotheses can be set on this research question. However, it is expected that similar school-level differences between teacher-given grades and students' attainment in standardized assessments as have been observed repeatedly in the basic school (Ouakrim-Soivio, 2013) will also be found at the upper secondary level.

3.2 Data and methods

The data to answer Research Questions 1 and 2 comprise the matriculation examination results of 131,089 matriculates (75,517 females, 55,572 males) between the years 2006 and 2009, and the 31,595 matriculates (18,243 females, 13,352 males) of spring 2012. For Question 3, additional data was collected in spring 2012 of the upper secondary course choices and grades of 1,997 students (1,147 females, 850 males) in a random sample of 35 schools.

For the matriculation examination, a mean score was calculated for each student based on all the exams the student sat for (e.g., Finnish, B-level mathematics, A-level English, C-level French, History, and Geography), using the numeral values of the different grades.⁶ For Research Question 3, a mean score was calculated for students' course attainments in upper secondary school (sum of courses attended and mean of grades awarded).⁷

Due to the exploratory nature of the study and the fractured data with each student only sitting for an average of 5.5 of the 39 exams available, the analyses will be limited to basic statistical methods. All analyses have been executed using SPSS 22.

4. Results

4.1 Research Question 1

It was hypothesized that as the students sitting for the different exams differ by general ability level, students sitting for exams in the more demanding subjects with more courses, especially A-level Mathematics and Physics, get in these exams

⁶ The scale is continuous between Approbatur = 2 points and Laudatur = 7 points (see Footnote 2) but discontinuous for Improbatur (failed) = 0 points.

⁷ The scale for grades runs from 4 (failed) to 10 (excellent).

lower grades than they get in the exams they share with a more heterogeneous examinee body.

The different exams and the share of students sitting for them are presented in Figure 1.⁸ Only 15 of the exams are sat yearly by more than 10 % of the examinees. As all education is offered in the two national languages of Finnish and Swedish (5.3 % of the population), the exam with most examinees is not Mother tongue and literature (the official name of the subject) but A-level English. The number of courses on which the different exams are based on, the number of examinees in 2006–2009, and the grade mean for each are presented in Table 1. There are clear differences in the grade means with the grades in B-level Swedish, A-level Mathematics, Physics, and Chemistry especially low, and the grades in C-level German especially high when compared to the grades in Finnish and A-level English which are sat by almost all students.

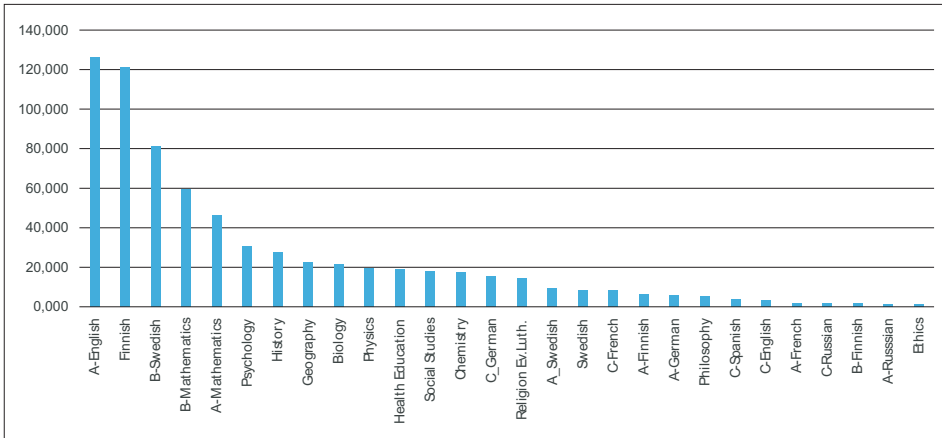
Table 1: Number of courses on which the matriculation examination exams are based on, number of examinees sitting for them in 2006–2009, and their grades (0, 2–7, see Footnote 2)

	Courses			N	Grade	
	Mandatory	Specialization	Total		M	SD
Mother tongue and literature	6	3	9	121,239	4.33	1.262
A-level language	6	2	8	126,327	4.35	1.465
B-level Swedish/Finnish	5	2	7	81,416	4.12	1.523
C-level language		8	8	15,610	4.46	1.497
A-level Math	10	3	13	46,365	4.11	1.654
B-level Math	6	2	8	59,225	4.20	1.581
Physics	1	7	8	19,515	4.14	1.544
History	4	4	7	27,557	4.33	1.470
Biology	2	3	5	21,691	4.23	1.502
Chemistry	1	4	5	17,316	4.15	1.616
Psychology	1	4	5	30,680	4.29	1.467
Religion/Ethics	3	2	5	14,622	4.27	1.480
Philosophy	1	3	4	5,280	4.25	1.485
Geography	2	2	4	22,507	4.31	1.457
Social Studies	2	2	4	17,726	4.23	1.523
Health Education	1	2	3	18,743	4.31	1.433

Note. The number of examinees and grades for Mother tongue is for Finnish (for Swedish $N = 8,508$, $M = 4.30$, $SD = 1.305$), those for B-level language (the other national language) for Swedish, those for A-level language for English (96.3 % of all students), those for C-level language German (the most common C-language exam), and those for Religion/Ethics for Lutheran religion (93.0 % of the examinees choosing one of the three exams of Lutheran or Orthodox religion or Ethics).

8 Nine exams with less than 1,000 examinees between 2006 and 2009 have been left out of the figure but not from the exam sum or grade mean calculated for each student.

Figure 1: Number of students sitting for the different exams, 2006–2009 ($N = 131,089$)



Note. Finnish and Swedish without an indication of level (A or B) refers to the language as mother tongue or language of instruction.

To test the hypothesis regarding students' grades in A-level Mathematics and Physics, found in Table 1 to be below the mean, in relation to the overall attainment of the students choosing them for their examination, we first examined how well the students earning the different grades in A- and B-level Mathematics had fared in the other exams they chose for their examination (Table 2). The mean grades for both A- and B-level Mathematics were seen in Table 1 to lag behind the mean grades for most other exams with the latter a little higher than the former. Table 2 shows, however, that the A-level Math students outperform the B-level Math students for all other subjects at all grade levels for their respective mathematics exams.

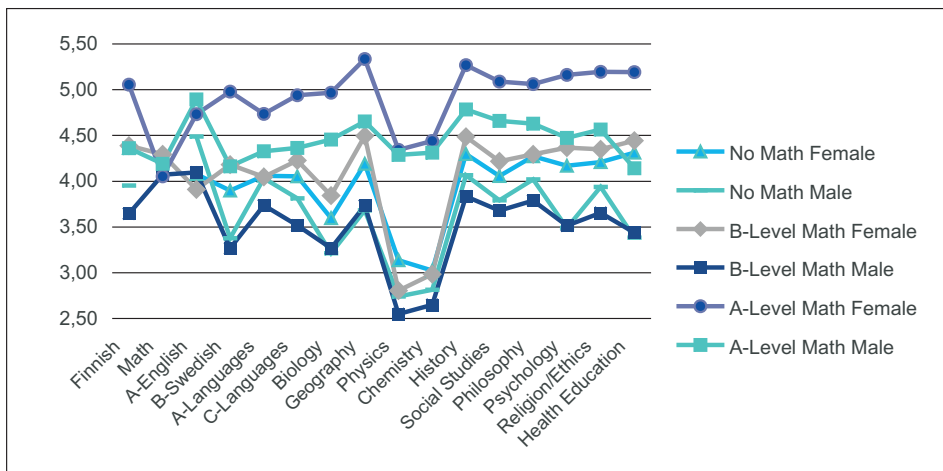
Table 2: Grade mean in exams other than mathematics according to the examinee's grade in A-level or B-level mathematics (2006–2009, $N = 46,365$ and $N = 59,225$, respectively)

	A-level	B-level
Laudatur	5.72	5.06
Eximia	5.27	4.54
Magna cum laude approbatur	4.86	4.13
Cum laude approbatur	4.47	3.77
Lubenter	4.15	3.48
Approbatus	3.79	3.20
Improbatur	3.99	3.62

To examine the hypotheses further, the student body was divided to three groups based on their choice of mathematics in the matriculation examination: A-level exam, B-level exam, or no mathematics exam⁹. The groups represent, respectively, 35 %, 45 %, and 20 % of the examinees. The ratio of male and female students in the groups (Figure 2) differ due to a combination of female students' bigger share of the examinees (57.6 %) and their greater propensity to shun mathematics. Thus, whereas 45.9 % of male but only 27.6 % of female students study A-level mathematics, the difference is much smaller among the examinees (55.1 % vs. 44.9 %). There is no difference in the share of male and female students studying B-level mathematics, so females are overrepresented in that group (57.7 % vs. 42.3 %) even if a much bigger share of female students also opt out of mathematics in the examination altogether (27.3 % vs. 9.2 %).

As can be seen in Figure 2, students sitting for the A-level mathematics exam outperform the other groups in all exams except for Mathematics and (for males) Health Education. Yet, their grades in the three subjects in which they are expected to be at their best, A-level Mathematics, Physics, and Chemistry, form clear *low* points in their attainment profile. This is especially true for A-level math females who otherwise so clearly outperform every other group.

Figure 2: Grades for all Finnish speaking matriculates in 2006–2009 and 2012 by gender and choice of mathematics exam (A-level, B-level, no math). Foreign languages other than A-level English have been combined into agglomerates of levels A and C due to the small number of examinees



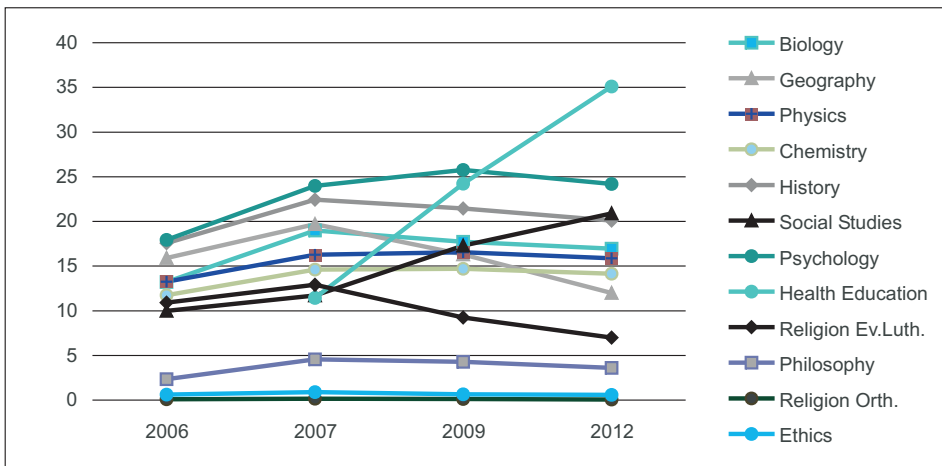
9 As only the exam in mother tongue is mandatory, students can omit mathematics from their examination even if they have to study at least six B-level courses of mathematics.

4.2 Research Question 2

It was hypothesized that due to students orienting themselves toward subjects and exams in which their chance for getting a better-than-expected grade would be greatest, there would emerge a trend especially for the weaker students to favor exams in subjects with fewer courses and a weaker examinee body.

Students' exam choices in 2006/2007¹⁰, 2009, and 2012 show that there have been quite dramatic changes in the popularity of the different exams since the reforms of 2005 and 2006. The most notable change regards the B-Level exam in Swedish for Finnish-speaking students which lost its mandatory status in 2005, with the share of examinees falling from 71.5 % in 2006 to 48.8 % in 2012. Among the natural science and socio-historical subjects which got their own exams in 2006, the addition of a separate exam for Health Education in 2007 with its mere three courses to study and revise for the exam attracted an immediate shooting star-ascent in popularity, causing many of the other exams to loose examinees (Figure 3). Even if the number of students studying A-level languages other than English and/or C-level languages is small, a similar loss of popularity can be seen in many of them as well with the examinees of A-level German declining from 16.1 % to 6.8 % between the years 2006 and 2012, and the examinees of A-level French from 8.1 % to 3.7 %.

Figure 3: Students' choices of exams in the Natural science/Socio-humanistic subjects in 2006, 2007, 2009 and 2012 (percentage of students sitting for the exam)



¹⁰ A separate exam for Health Education was adopted only one year after the exams in the other science/socio-humanistic subjects.

4.3 Research Question 3

It was hypothesized that there are between-school differences in the relation of students' teacher-given course grades and their attainment in the different exams in the matriculation examination.

Students are selected to the upper secondary school of their choice based on their grade point average (GPA) in ninth grade, resulting in clear differences between schools in terms of students' prior achievement level, especially in the bigger cities. This can be expected to cause between-school differences in students' subject and course choices at school, and reflected in the share of students choosing the different exams in the matriculation examination. This, in turn, can affect the relation between course grades and exam results. Accordingly, we will first look shortly at between-school differences in students' exam choices. As students' attainment in the different exams of the matriculation examination is clearly related to their choice of mathematics exam (see Figure 2) the most salient difference between schools regarding students' choices is expected to be the difference in the share of students in the three groups of A-level, B-level, and no Math. In the schools with more than 25 matriculates included in this analysis ($N = 338$), the share of students sitting for the A-level Mathematics exam varied from 0 to 74 %. The overall differences were not very big, however, with the variance in the share of students sitting for the A-level slightly bigger than that for B-level ($\eta^2 = .08$ vs. $\eta^2 = .05$). Neither was related to school size in terms of the total number of matriculates. A clearer difference between schools ($\eta^2 = .13$) could be found in the B-level Swedish for Finnish students, the most controversial subject within the Finnish education system (see Mandatory Swedish, n.d.) and shunned especially by male students. Some fluctuation was to be expected due to differences in the linguistic characteristic of the municipalities and in 10 % of the 338 schools with more than 25 Finnish-speaking matriculates there were no examinees in B-Swedish.

The 2012 sample data allows looking simultaneously at students' course choices and success at school, and their exam choices and success in the matriculation examination. Due to attrition, big schools (101 or more matriculates per year) are somewhat underrepresented in the sample (14 % vs. 25 %) but the distributions of exam choices and attainment are in accord with the whole 2012 matriculate body both for gender and for mathematics choice.

The correlation between students' matriculation examination grade and their attainment in the respective subject at school (mean of all courses, calculated only for the students taking the exam and expected to have passed all the required courses) varied between $r = .49$ and $r = .80$ (Religion/Ethics and A-level English, respectively) with a mean of $r = .69$ (Table 3). The differences may reflect differences in the reliabilities of the different exams, differences in the way the exams cover the respective syllabi, between-school differences in school grading, or the relatively small and selected groups sitting for the different socio-humanistic and natural science exams in the sampled schools. The correlation between the mean

for the whole matriculation examination and upper secondary GPA was relatively high at $r = .79$ with no gender difference.

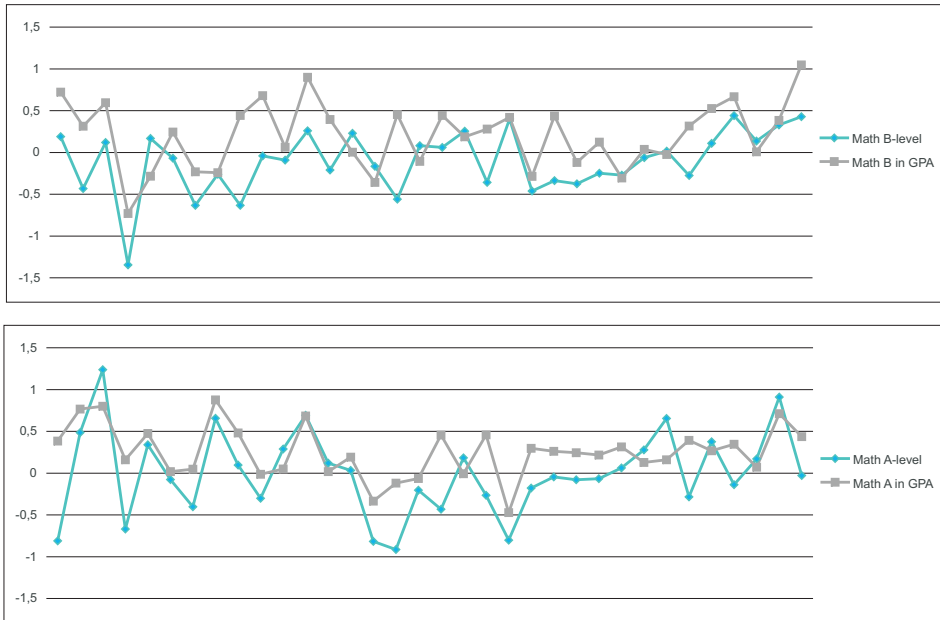
Table 3: Correlations between students' final school report grade and matriculation exam grade in the respective subject

	<i>r</i>
Finnish	.710
A-English	.801
B-Swedish	.791
Math Advanced	.737
Math Basic	.698
Physics	.719
History	.714
Biology	.704
Chemistry	.662
Psychology	.674
Religion/Ethics	.487
Geography	.647
Social Studies	.729
Health Education	.636
GPA/Mean	.791

Note. Philosophy has been left out due to the small number of examinees in this sample of 2,000 matriculates.

Differences between schools were statistically significant for most subjects both for the matriculation examination results and for school marks ($p \leq .001$, effect sizes between $\eta^2 = .073$ and $\eta^2 = .215$). Students' (standardized) grades for A- and B-level Mathematics in GPA and in the matriculation examination show considerable between-school fluctuation for both (Figure 4). As the students not sitting for the mathematics exam don't have comparable school grade means due to the smaller number of courses they have studied, it is hard to assess their role in the discrepancies between the two scores.

Figure 4: Students' grades in A-level (advanced) and B-level (basic) mathematics in the matriculation examination and in GPA (school grades) by school (sample of 35 schools, standardized values)



A disconcerting additional finding is that the school grades varied considerably between subjects. While the mean for B-level Mathematics was 6.98 on the scale of 4 (failed) to 10 (excellent), the mean for Health Education was 7.94 which can be suspected to further affect students' course and exam choices. There was a clear overall tendency for the school grades of the more demanding subjects with more courses and more able students to fall well below the mean for all subjects (7.65) while the more popular and easier subjects with less courses fell above it.

5. Discussion

Two key reforms regarding the Finnish matriculation exam took place during the first decade of the 2000s. The first increased exam choice by waiving the mandatory status in the examination of the second national language (Swedish for the Finnish-speaking students, Finnish for the Swedish-speaking) by making it just one of the elective exams. The other reform increased exam choice by dividing the earlier incorporated exam for all socio-humanistic and natural science subjects to ten different exams. The aim of this reform was to reinforce the weight of these subjects in the examination and in the school syllabus. As a consequence of these two reforms, the earlier fairly homogenous examination was replaced with one where only the exam in the language of instruction (Finnish/Swedish) is mandatory, in

addition to which the examinee has to choose three other exams from the groups of (a) mathematics, (b) one foreign language, (c) the second national language, and (d) one exam in natural and socio-humanistic sciences. In addition to these, the examinee is free to choose auxiliary exams in foreign languages and in natural and socio-humanistic sciences.

Following an earlier reform of upper secondary education introducing a class-level-free syllabus with increased choice, the examination reforms can be seen as an effort to accommodate the increasing heterogeneity of the student population since the adoption of the comprehensive school in the 1970s. The present study is the first to explore the impact of these reforms on the comparability of the matriculation examination results, a key issue regarding the high stakes of the examination due to the use of its results in selecting student to tertiary education.

The fairly homogenous earlier examination with less choice relied on the relative comparability of grades between the different exams and across years due to the two level grading comprising a Gaussian distribution of final grades based on a preliminary criterion-based scoring of the different subject-specific exams (cf. Howie, Long, Sherman, & Venter, 2008). As the present study shows, however, the Gaussian distribution has become a hinder for comparability due to the fractured student body of varying academic ability sitting for the different exams. This should not come as a surprise regarding the relation of the close to forty different exams on offer and the average 5.5 exams that the students include in their examination. The comparability of the grades have earlier been examined by subject across years (Mehtäläinen & Välijärvi, 2013) while the present study is the first to systematically look at the grades in relation to the overall success of the students sitting for them. Admittedly, this presupposes the existence of a common factor for academic ability. However, the presupposition is supported by the high correlation between students' attainment in the different exams despite the apparent distortion caused by the differences in the levels of the exams in terms of the number of courses they cover.

The results confirm the original unwritten hypotheses behind the study that many students choose exams offering optimized value for their ability or willingness to exert effort, and that this optimization is encouraged by the Gaussian distribution of grades with students of differing ability level attaining the same grade due to differing exam choices. While helping the weaker students to attain in some exams grades which surpass their general level of competence, this grade-inequity devalues the proficiency of the more able students who compete against each other for grades in the more demanding subjects. This is seen most clearly in the grades of the female students of A-level mathematics, the highest achieving group of all, whose grades in mathematics, physics and chemistry – the three subjects they focus specifically on – fall behind their grades in all the other exams they choose to their examination.

The results also show that exams where getting a good grade is relatively easier have increased in popularity year after year, even if the data does not allow drawing any conclusions as to the direction of the causality: Is the small number of

courses the driving force behind the exam choice or does the possibility of avoiding competition with the more able students who seldom choose Health education into their exams, lead the course choice?

Even if there are cultural factors (see Mandatory Swedish, n.d.) responsible for the decrease of the popularity of the B-level Swedish exam since the renouncement of its mandatory status (students still have to study the five mandatory courses at school, and their course grades will be included in the final report card), students also weigh on the probability of achieving in the exam a grade that would justly reflect their competence in relation to other students. It is evident that as the number of students sitting for the exam decreases, the attrition grows more rapidly at the weaker end and ever better students have to content themselves with a grade lower than their actual level of ability, creating a vicious circle further decreasing students' desire to choose the exam. The same inflationary development applies to all exams with better than average examinees, be it A-level Mathematics, Physics, Chemistry, or an A-level language other than English. In all of these, only the very best students can confidently expect a grade that will duly reflect their ability vis-à-vis the whole student body. The development is contrary to the exams in the less demanding subjects where the less able students can expect to get grades that exceed their ability vis-à-vis their better achieving peers who do not value such exams enough to include them in their examination.

All in all, the results show that the goal of the recent reforms to better serve students by increased choice in the syllabus and in the matriculation exam has backfired in many ways. In schools, the increased choice has created a group of students whose education no more offers the officially stated well-rounded education, covering adequately mathematics, languages, and socio-humanistic and natural sciences. However, it might also be that the syllabus and the structure of the studies requiring a fair amount of self-regulation and planning skills just is too demanding for some of the students entering the general upper secondary track after the comprehensive school.

It is evident that the increased choice has led to the more able students competing against each other for grades in the more demanding subjects while the weaker students receive the (nominally) same grades in the less demanding exams. Even if there is no earlier research on this phenomenon, the results are clearly no secret for students as can be seen in the changing popularity of the different exams. As it is, the reforms seem to have led many students – especially those of weaker-than-average attainment – to aim at getting as good grades as possible with as little work as possible. The concern is for those students who outperform their peers at school but have to resign to their lower-than-expected grades in the more demanding subjects.

Through the results, the study reveals John Bishop's construct for CBEEES to lack or to include just as an unwritten supposition one key premise, namely, that to fully advance the achievement of all students, an exam system should have all students sit for the same exams at least in the subjects deemed critical for that goal.

The findings have a special urgency in Finland vis-à-vis plans to reinforce the role of the matriculation results and/or the school grades in admission to university. The findings also emphasize the need to look for ways to solve the dilemma between the goals of equity and fairness in exit exams in other education systems. It is worthwhile to return to Coe's and his colleagues' (2008) conclusion regarding their findings on the relative difficulty of the examinations of different subjects in the British A-level and GCSE examinations: "Given the evidence about the relative difficulties of different subjects, we believe there are three possible options for policy: to leave things as they are, to make grades statistically comparable, or to adjust them for specific uses" (p. 3).

Having been informed of the preliminary results of the present study, the Finnish Matriculation Examination Board has introduced a statistical method for increased grade comparability based on the average of standardized scores for the whole examinee body instead of separate exams (FMEB, 2015b; cf. Marjanen, 2015).

Acknowledgements

The research was funded by the Finnish Ministry of Education and Culture. We are most grateful to the referees and the editor(s) of JERO for their helpful comments.

References

- Amrein, A. L., & Berliner, D. C. (2002). High-Stakes testing, uncertainty, and students learning. *Education Policy Analysis Archives*, 10(18), 1–74.
- Béguin, A. A. (2000). *Robustness of equating high-stakes tests*. Thesis, University of Twente, Enschede, Netherlands.
- Bishop, J. (1998). The effect of curriculum-based external exit exam systems on student achievement. *Journal of Economic Education*, 29(2), 172–182.
- Bishop, J. H., Mañe, F., & Bishop, M. (2001). How external exit exams spur achievement. *Educational Leadership*, 59(1), 58–63.
- Coe, R. (2008). Comparability of GCSE examinations in different subjects: An application of the Rasch model. *Oxford Review of Education*, 34(5), 609–636.
- Coe, R., Searle, J., Barmby, P., Jones, K., & Higgins, S. (2008). *Relative difficulty of examinations in different subjects*. Durham, United Kingdom: CEM Centre, Durham University
- Csapó, B. (2009). The second stage of public education and the Matura. In K. Fazekas, J. Köllő & J. Varga (Eds.), *Green Book for the renewal of public education in Hungary* (pp. 81–104). Budapest, Hungary: Ecostat.
- EF – Education in Finland. (2012). *Finnish education in a nutshell. Education in Finland*. Retrieved from http://www.oph.fi/download/146428_Finnish_Education_in_a_Nutshell.pdf
- FMEB – Finnish Matriculation Examination Board. (2015a). *The Finnish matriculation examination*. Retrieved from <https://www.ylioppilastutkinto.fi/fi/english>

- FMEB – Finnish Matriculation Examination Board. (2015b). *Improving the comparability of grades*. Retrieved from https://www.ylioppilastutkinto.fi/images/sivuston_tiedostot/Ylioppilastutkinto/Improving%20the%20comparability.pdf
- FNBE – Finnish National Board of Education. (2003). *National core curriculum for upper secondary schools 2003* (Regulation 33/011/2003). Helsinki, Finland: Finnish National Board of Education.
- Howie, S. J., Long, C., Sherman, V., & Venter, E. (2008). *The role of IRT in selected examination systems*. Pretoria, South Africa: Umalusi Council of Quality Assurance in General and Further Education and Training.
- Jürges, H., Schneider, K., & Büchel, F. (2003). *The effects of central examinations on student achievement: Quasi-Experimental evidence from TIMSS Germany* (Ifo Working Paper No. 939). Retrieved from CESifo Group Munich website: http://www.cesifo-group.de/portal/page/portal/DocBase_Content/WP/WP-CESifo_Working_Papers/wp-cesifo-2003/wp-cesifo-2003-05/cesifo_wp939.pdf
- Jürges, H., Schneider, K., Senkbeil, M., & Carstensen, C. H. (2012). Assessment drives learning: The effect of central exit exams on curricular knowledge and mathematical literacy. *Economics of Education Review*, 31(1), 56–65.
- Kaarninen, M., & Kaarninen, P. (2002). *Sivistyksen portti. Ylioppilastutkinnon historia* [A gate to higher education. History of the matriculation examination]. Keuruu, Finland: Otava.
- Kahnert, J., Eickelmann, B., Bos, W., & Endberg, M. (2012, December). *Mathematical competencies in high school exit examinations and in TIMSS/III – Validation of high stakes-testing through an international large scale assessment instrument*. Paper presented at the WERA and Joint AARE APERA International Conference, Sydney, Australia.
- Klein, E. D. (2013). *Statewide exit exams, governance, and school development: An international comparison*. Münster, Germany: Waxmann
- Klein, E. D., & van Ackeren, I. (2011). Challenges and problems for research in the field of statewide exams. A stock taking of differing procedures and standardization levels. *Studies in Educational Evaluation*, 37(4), 180–188.
- Lukiolaki. (1998). *Lukiolaki (1998/629)*. [General Upper Secondary Schools Act]. Retrieved from <http://www.finlex.fi/fi/laki/ajantasa/1998/19980629#L4P18>
- Maag Merki, K. (2011). Effects of the implementation of state-wide exit exams on students' self-regulated learning. *Studies in Educational Evaluation*, 37(4), 196–205.
- Mandatory Swedish (n.d.). *Mandatory Swedish*. Retrieved from https://en.wikipedia.org/wiki/Mandatory_Swedish.
- Marjanen, J. (2015). Ylioppilasarvosanojen vertailukelpoisuusongelma ja SYK-menettelmän edellytykset sen ratkaisemiseen. [The problem of grade incomparability in the matriculation examination and the Standardized Marks Average-method as a possible solution]. *Kasvatus*, 2015(4), 317–333.
- Mehtäläinen, J., & Välijärvi, J. (2013). *Ylioppilaskokeiden arvosanojen vertailtavuus eri aineissa vuosina 2007–2011* [Comparability of matriculation examination grades across years in 2007–2011]. Tutkimusselosteita 46. Jyväskylä, Finland: Koulutuksen tutkimuslaitos, University of Jyväskylä.
- Noah, H. J., & Eckstein, M. A. (Eds.) (1992). *Examinations. Comparative and international studies*. Oxford, United Kingdom: Pergamon Press.
- Ouakrim-Soivio, N. (2013). *Toimivatko päättöarvioinnin kriteerit? Oppilaiden saamat arvosanat ja Opetushallituksen oppimistulosten seuranta-arviointi koulujen välisten osaamiseröjen mittareina*. [Do the national criteria for students' final grades work? Teachers' grading and national assessments as indicators for between-school differences]. Helsinki, Finland: Opetushallitus.
- Reardon, S. F., Atteberry, A., Arshan, N., & Kurlander, M. (2009). *Effects of the California high school exit exam on student persistence, achievement, and graduation* (Working paper 2009–12). Retrieved from Institute for Research on

- Education Policy & Practice website: http://web.stanford.edu/group/cepa/workingpapers/WORKING_PAPER_2009_12.pdf
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, 25(1), 54–67.
- VNK – Valtioneuvoston kanslia [Prime Minister’s office]. (2011). *Pääministeri Jyrki Kataisen hallituksen ohjelma*. [Programme of Prime Minister Jyrki Katainen’s Government]. Helsinki, Finland: VNK.
- Vuorio-Lehti, M. (2006). *Valkolakin viesti. Ylioppilaskeskustelu Suomessa toisen maailmasodan jälkeen*. [The message of the White Cap. The matriculation discourse in Finland after the Second World War]. Turun yliopiston julkaisuja. Annales Universitatis Turkuensis. Sarja – Ser. C Osa – Tom. 253 Scripta Lingua Fennica Edita. Turku, Finland: University of Turku.
- Waldow, F. (2009). What PISA did and did not do: Germany after the ‘PISA-shock’. *European Educational Research Journal*, 8(3), 476–483.