The Effect of the Language of Instruction on Academic Performance

Juliana Bernhofer[1] and Mirco Tonin[2]

[1,2]Faculty of Economics and Management, Free University of Bozen-Bolzano

January 31, 2020

Abstract

With the growing international mobility of students, it is becoming more and more common to study in a language other than the native one. To explore the impact of the language of instruction on grades, we exploit the unique features of a comprehensive dataset provided by the Free University of Bozen-Bolzano. By analyzing data from an officially trilingual university which is located in a multilingual institutional framework, we are able to tackle the problem of self-selection typical to these kind of analyses. Our results suggest that even though there are no *ex ante* asymmetries in terms of grades between speakers of different native languages, taking an exam in a second language leads to a loss in grade points of approximately 8.5%. We also look into the effort allocation of non-native learners by analyzing the exam success rate and bring evidence to some strategic elusive behavior of students in terms of exam language portfolio composition. Finally, we analyze the effectiveness of different second language acquisition tools and provide policy recommendations for both academic and professional purposes.

Keywords: Language and Economic Behavior, Education, Multilingual Education, Higher Education

JEL Classification: I21,J24,Z13

*E-mail addresses: juliana.bernhofer@unibz.it, mirco.tonin@unibz.it

1 Introduction

In December 2017, a special session of the *Salzburg Global Seminar Series* has produced *The Salzburg Statement for a Multilingual World* (Regester and Norton, 2018). The meeting brought together experts from the private and the academic sector and looked specifically at multilingualism and its importance to dynamic and entrepreneurial societies, as well as the importance of language rights for minority groups around the world. The issue was assessed from a number of different perspectives, such as policy making, language rights, social cohesion and language teaching. Policy recommendations specifically address the need of *"utilizing insights from educational and cognitive research for mother tongue and other-tongue learning"*. Besides its policy relevance, our research also has a strong interdisciplinary link with the area of psycho- and neurolinguistics and in particular the literature covering Second Language Acquisition (SLA).

Although it is recognized that multilingual education is a tool to decrease inequalities in the labor market within and between regions and enhances cultural awareness (Isphording, 2015), little is known about how languages differ in their cognitive influences on behavior and on learning processes and how such asymmetries affect the efficiency and efficacy of multilingual education programs. Imbalanced outcomes of multilingual learning processes could be speaker-driven, when they arise from the grammatical structure of the language in question, and/or exam-driven, when some courses feature a more domain-specific terminology than others. In order to address issues of inequality it is important to tackle asymmetries that may arise during the process of language acquisition and application.

Our sample of analysis are students from the Free University of Bozen-Bolzano (the University henceforth) in South Tyrol (Bolzano - Alto Adige), an autonomous province in Northern Italy with Bozen-Bolzano as its capital. The formerly Austrian-Hungarian county was annexed to the Kingdom of Italy in 1919 and is today officially trilingual (German, Italian and Ladin for some municipalities). During the years of Fascism, South Tyrol was subject to interventions of italianization, with a total of 32 measures, as for example bans on the German language, subsidized immigration from other Italian regions and the choice between the Italian and the Austrian citizenship with the expulsion of those who opted for the latter. Today, around 62% speak German as their first language, 23,4% speak Italian and the remainder Ladin or other languages. The Free University of Bozen-Bolzano was founded in 1997 with the explicit mission of trilingual eduction

(Italian, German and English).

For most of the students, one of the three languages is their native (L1) language, while they have varying degrees of knowledge of the other two (L2) secondary languages. This allows us to observe the performance of the very same student as he or she learns in the native language and in other languages and to compare it with the performance of course mates with a different linguistic background. This feature is crucial for the study. Our research questions could not be addressed using e.g. English and foreign students at a UK university because, due to self-selection, there is no clear comparison group (i.e. it may well be the case that, on average, foreign students attending a UK university may be of higher ability than British students and this would mask any language disadvantage they may suffer when studying in a foreign language).

Our main research questions are the following:

1. Are there, on average, differences and asymmetries in academic performance based on the linguistic background of the students and/or the language of the exam?

2. Do such linguistic differences disappear over time?

3. Which are the most appropriate tools to tackle linguistic differences (internal linguistic exams vs. external exams vs. periods abroad)?

4. Do students exhibit strategic behavior to elude the trilingual study plan policy?

The working research hypothesis is that, while there may be an initial disadvantage in studying in a language other than the native one, this may become less and less important and eventually disappears as students progress with their course of study. By analyzing subsamples of students, this study also allows to identify heterogeneity across groups, thus making it possible to design specific interventions to support individual study careers and to structure the study programs for optimal learning. This study also provides evidence on the relative benefit of different models of language acquisition (standardized language certificates, internal exams, periods spent abroad) and draws attention to the fact that even well-intentioned language policies might fail to be fully effective if the population in question is assumed to be intrinsically motivated.

2 Literature Review

2.1 The Economics of Education

A large and vibrant body of research within the economic and non-economic literature examines the determinants of educational outcomes on one side, and the impact of education on life outcomes, such as wages, career prospects, political participation and social status on the other. One of the indisputable pioneers of the field of economics of education was Schultz (1960), who introduced the concept of *Human Capital* as the output of the educational "investment in man". The treatise of Becker (1964) took the concept of investment in Human Capital and the centrality of education alongside on-the-job training, medical care, and migration further, by elaborating the idea of the return to investment in education in relationship to the human lifespan.

The economic approach to education in considering rates of return and trade-offs in the analysis of ex-post effects on labor market outcomes eventually led to the necessity of looking into the determinants of educational achievements and the development of education production functions (EPF henceforth) to measure the effect of various schooling inputs on test score outcomes for children in school age (Todd and Wolpin, 2003). As of today, a number of determinants of educational outcomes are known to play a role with different levels of impact in primary, secondary and tertiary education. Much work has been done on Early Childhood Development (ECD) and on the EPF of schooling since the Beckerian prelude on Human Capital, whereas higher education gained relevance more recently, also as a result of societal changes during the last three decades which led to an increase in the demand for tertiary education.

Whereas parental factors and other household variables are shown to play a central role in the academic achievement of school-aged students (Coleman, 1966), in the analysis of post-secondary outcomes, more weight is put on the individual characteristics of the student. This is also due to the fact that the aforementioned parental and environmental factors are partly captured by the high school grade, which is considered to be one of most relevant predictors of academic performance in higher education (Griffith and Rask, 2014; Porchea et al., 2017). Other factors are likely to have separate effects on all educational levels, such as the broader concept of identity and self-image (Akerlof and Kranton, 2003), the quality of instructors and of the educational environment (Chetty et al., 2014b,a; De Paola, 2009; Hoffmann and Oreopoulos, 2009), peer-effects (Oosterbeek and Van Ewijk,

2014; Griffith and Rask, 2014) and gender and ethnicity (Thiele et al., 2016). Within the non-economic literature, greater attention is put on motivational and behavioral factors, such as educational persistence, self-confidence, commitment and social activity (Robbins et al., 2006; Tinto, 1997).

2.2 Multilingual Learning

The phenomenon of bilingual and multilingual learning within the economic literature has been mainly analyzed in relationship with labor market outcomes (Cappellari and Di Paolo, 2018; Saiz and Zoido, 2005), but the strand generally falls short of considering educational efficiency and effectiveness of multilingual learning environments. How labor market outcomes differ based on language and culture is described for example by Brügger et al. (2009), who analyze the population along the Germanic-Romance language border in Switzerland. They find an unemployment duration which is 20% higher within the Romance-speaking border community compared to German-speakers.

Related theories and findings from cognitive neuroscience and psycholinguistics provide some interesting results on how language affects cognitive aspects and reasoning. Venkatraman et al. (2006); Grabner et al. (2012) for example use functional magnetic resonance imaging (fMRI) on bilingual subjects to show how language-switching costs could impact on the efficacy of bilingual learning. Language-switching costs apply to those courses in which the student has to mentally retrieve information that was previously acquired in a different language to the language of the exam, for example math skills during high school needed to pass an exam in microeconomics. The mere cost of switching from one language to another, even when the student is perfectly bilingual, represents an additional cognitive effort compared to a situation in which the language of acquisition does not differ from the language of learning. Furthermore, Abutalebi and Clahsen (2017) provides an overview of the current state-of-art in the psycholinguistics of bilingual learning and a recent article of Cunnings (2017) gives a new interpretation of the mechanism underlying native (L1) and non-native(L2) language processing, which differs from the acclaimed shallow structure hypothesis (SSH) (Clahsen and Felser, 2006) according to which differences in native and non-native processing are due to a mechanism of shallow parsing. Rather than qualitative differences and in line with the idea of the aforementioned language-switching costs, he interprets L1/L2 differences as a consequence of retrieval inferences during sentence processing from memory for non-native speakers.

Going back to the economic concept of education production functions, with this work we shorten the gap between the economic and non-economic literature by integrating linguistic aspects into the baseline educational production function as it is described by (Hanushek, 1995, 1979; Cappellari et al., 2012) and assess whether and to what extent non-native learning impacts on academic performance.

3 Theoretical Framework

A simple model of learning effort

From a general point of view, students are maximizing their utility given by $U(L, G)$ where L represents the amount of hours dedicated to leisure and G the grades in terms of GPA achieved thanks to the time invested in studying. If we assume, that their preferences follow a Cobb-Douglas specification, then leisure and GPA translate as follows into utility:

$$U(L, G) = L^{\alpha} G^{\beta}, \quad \alpha + \beta = 1 \tag{1}$$

where α and β represent leisure's and GPA's share of utility. Our simplified model assumes by construction that effort exerts a direct negative effect on the student's utility function, but there are two channels of effort: the direct negative effect given by the lower availability of leisure, but also a positive effect induced by higher grades, which are increasing in the level of effort. The overall level of GPA is given by the average of the single grades achieved by the student, such that $G = \frac{\sum_{i=1}^{n} g_i}{n}$ and

$$g_i = A_i e_i^{\gamma}, \quad A > 0, 0 < \gamma < 1, i = 1, ..., n \tag{2}$$

The technology parameter A_i determines the extent to which effort affects grades and can be generally interpreted as the *study technology*, which broadly stands for the student's skill set he or she acquired previously to the effort allocation decision. However, we can interpret the technology also as a parameter able to distinguish exams taken in the student's mother tongue and exams taken in L2. The subscript i allows for exam-specific technologies, a type of flexibility we will need at a later point of our model specification. The exponent γ also describes the impact of effort on grades, but in marginal terms by affecting the curvature of the production function. By constraining γ between 0 and 1, we assume decreasing returns to scale. Assuming a rule such that there have to be both, exams in the student's mother tongue (L1) and exams in a second language (L2) in equal

proportions in the study plan, in order to maximize the GPA we find the effort levels e_i, with $i = 1, 2$, that maximize grades for some level of overall effort $k = \sum_{i=1}^{n} e_i$.

$$\begin{aligned} \max \quad & G = \frac{g_1 + g_2}{2} \\ \text{s.t.} \quad & g_1 = A_1 e_1^{\gamma} \\ & g_2 = A_2 e_2^{\gamma} \\ & e_1 + e_2 = k \end{aligned} \tag{3}$$

We first find the optimizing level of effort to be invested in the L1 exam, ${e_1}^*$ by substituting all constraints into the objective function and by putting its first derivative equal to zero.

$$\begin{aligned} & \max_{e_1} G = \frac{A_1 e_1^{\gamma} + A_2 (k - e_1)^{\gamma}}{2} \\ & \frac{dG}{de_1} = A_1 {e_1}^{\gamma - 1} - A_2 (k - e_1)^{\gamma - 1} = 0 \end{aligned} \tag{4}$$

with the optimal level of e_1 being

$${e_1}^* = k \frac{\left[\frac{A_1}{A_2}\right]^{\frac{1}{1-\gamma}}}{\left[\frac{A_1}{A_2}\right]^{\frac{1}{1-\gamma}} + 1} \tag{5}$$

By substituting (5) into the constraint $e_1 + e_2 = k$, we obtain:

$${e_2}^* = k \frac{1}{\left[\frac{A_1}{A_2}\right]^{\frac{1}{1-\gamma}} + 1} \tag{6}$$

It is easy to see that when we have the same production technology for all exams $A = A_i$, the optimal levels of effort become $e_1 = e_2 = k/2$. However, if we allow for different production technologies for L1 and L2 exams, such that $A_1 > A_2$, we are able to model the fact that the grade production process in a student's native language has superior technology. By assuming decreasing returns to scale $0 < \gamma < 1$, it follows that the student optimally invests a higher level effort into the exam in L1 with $e_1^* > e_2^*$.

By simplifying $\left[\frac{A_1}{A_2}\right]^{\frac{1}{1-\gamma}} = z$, our student's optimal level of grade production results as follows:

$$G^*_{A_1,A_2} = \frac{A_1\left[k\frac{z}{1+z}\right]^\gamma + A_2\left[k\frac{1}{1+z}\right]^\gamma}{2} \tag{7}$$

We can now find the optimal level of overall effort k^* by substituting (7) into the student's utility function we initially defined in (1).

$$U(G^*_{A_1,A_2}) = (1-k)^\alpha \left[\frac{1}{2}\right]^\beta \left(A_1 k^\gamma \left[\frac{z}{1+z}\right]^\gamma + A_2 k^\gamma \left[\frac{1}{1+z}\right]^\gamma\right)^\beta \tag{8}$$

By working with the first-order condition of (8), we find that

$$\begin{aligned}
&\frac{k}{1-k} = \frac{\beta\gamma}{\alpha} \\
&1-k = L \\
&k = \frac{\beta\gamma}{\alpha} * L \\
&L = 1 - \frac{\beta\gamma}{\alpha} * L \\
&L^* = \frac{1}{1+\frac{\beta\gamma}{\alpha}} = \frac{1}{1+c} \\
&k^* = \frac{c}{1+c}
\end{aligned} \tag{9}$$

It is interesting to note that the level of overall effort does not depend on the production technology A_i, but solely on the parameters defining production efficiency and personal preferences. If we assume on the other hand that the student studies in an exclusively native linguistic environment, such that for all exams $A_2 = A_1$, we already found that the optimal choice is $e_1 = e_2 = k/2$, which results in the following optimal grade production function

$$G^*_{A_1} = A_1\left(\frac{k}{2}\right)^\gamma \tag{10}$$

If we use the optimal values for effort in the monolingual context to calculate the grade output in a bilingual environment, we have that

$$G_{A_1,A_2}\left(\frac{k}{2}\right) = \frac{A_1\left(\frac{k}{2}\right)^\gamma + A_2\left(\frac{k}{2}\right)^\gamma}{2} \tag{11}$$

The difference between G_{A_1} and $G_{A_1,A_2}\left(\frac{k}{2}\right)$ gives us the *technical component* of the differences in grades between bilingual and monolingual study plans and is equal to

$$G^*_{A_1} - G_{A_1,A_2} = \Delta_{TC}G = \frac{A_1\left(\frac{k}{2}\right)^\gamma - A_2\left(\frac{k}{2}\right)^\gamma}{2} \tag{12}$$

The overall difference between grades is given by the difference between the optimal levels of $G^*_{A_1,A_2}$ and $G^*_{A_1}$ of (7) and (10) and includes the technical component in (12) which is due to the different production technologies alongside a behavioral mechanism shifting the level of effort for L1 exams upwards.

$$G^*_{A_1} - G^*_{A_1,A_2} = \Delta_{TOT}G = A_1\left(\frac{k}{2}\right)^\gamma - \left[\frac{A_1\left(\frac{z}{1+z}\right)^\gamma k^\gamma + A_2\left(\frac{1}{1+z}\right)^\gamma k^\gamma}{2}\right] \tag{13}$$

If for example university rules impose both language groups to be a mandatory part of the study plan in equal proportions, the student has an incentive to work on the technology parameter A_2 in order to increase the overall utility level. She or he could do so by attending additional targeted tutorial sessions that could increase the efficiency of non-native learning. Also, intrinsic motivation and/or better labor market outcomes are additional parameters with the potential to increase the overall utility of the multilingual student.

By comparing the utility function of a student who only takes exams in L1 with the utility function of a student who has to have both, exams in L1 and L2 in her or his study plan, we find the parametric cost C of studying in a non-native environment.

$$\begin{aligned} &\Delta U = U_{A_1} - U_{A_1,A_2} = C \\ &C = (1-k)^\alpha \left[\left[A_1\left(\frac{k}{2}\right)^\gamma\right]^\beta - \left[\frac{A_1 {e_1}^\gamma + A_2 e_2^\gamma}{2}\right]^\beta\right] \end{aligned} \tag{14}$$

The higher the difference between production technologies A_1 and A_2, the higher will be the cost of having to dedicated effort to L2 learning. The representation of the problem in terms of costs becomes even more relevant if we assume that potential students have a myopic decision model, which does not account for better labor market outcomes thanks to the acquired language skills. Also, students which are already enrolled have an incentive to invest in strategies aimed at eluding the multilingual study plan.

4 Institutional Framework

Our full database contains information about more than 15'000 students enrolled at the Free University of Bolzano which was founded in 1997. The University is located in

Bolzano, the capital city of the province of South Tyrol in northern Italy and actively promotes multilingual education as part of their mission and mandatory study plans. Overall, the University has five faculties (Economics and Management, Education, Computer Science, Science and Technology, Design and Art) and offers courses in German, Italian and English. The Faculty of Education has a more vertical structure in terms of multilingualism, as its main aim is to train primary and secondary school teachers and therefore a student's study plan tends to focus on one of the region's main teaching languages which are German, Italian or Ladin. Since 2011 the University started to register with accuracy type and date of language certificates and exams of the students, as well as the teaching language of the courses.

5 Data Description

Data is anonymous and for each student ID we can explore the grades for each exam, as well as socio-demographic information such as their birth province, gender, age and high school grade. Starting from the year 2011 up to today we also have reliable information about the type and date of certificates or exams as a proof of the student's language level and are thus able to create a path of language acquisition to measure its impact on grades. Besides the student database, we also have information on the teaching language of the courses, the instructor's linguistic and socio-demographic characteristics and his or her academic role in the year of teaching.

After careful consideration, we decided to include only the courses of the Faculties of Economics and Management, Computer Science, Science and Technology and Design and Art into our analysis. The exclusion of the students enrolled in the Faculty of Education is due to the fact that the student and exam compositions are characterized by a strong bias towards German in terms of both mother tongue and exam languages. For a similar reason, we excluded 18 Master's degree programs (17.7% of the faculty-corrected sample) as they are either double degrees with other Italian university or are not trilingual in their languages of instruction. Furthermore, we also excluded incoming exchange students and students with other irregular enrollment status which account for another 11.7% of the sample. As a result, our final dataset contains 6'595 students for a total of 161'117 exams. Each student is identified by an individual student ID and an enrollment number for the chosen study course which allows us to control for individual effects in our regression analysis and to identify the 161 students who have been enrolled in two separate programs

with the University of Bolzano over their academic career.

Linguistic variables. As the University does not collect information on the students' native language, we assigned either English, Italian or German as native[1] language by proceeding as follows: those who had a high school degree (Abitur/Maturità/High School) in Italian, German and/or English as proof of their language level at the moment of enrollment were considered natives in that language[2]. The remainder were assigned with the language of their country of origin, if present and if the language was spoken by more than 80% of the resident population[3]. For those who had no information on the high school language and the language of the country of origin, we considered the chosen language of communication at enrollment. This is a choice students, instructors and personnel are given at the moment of registration with the system of the University, but was unfortunately not available for all subjects in our database. For the aforementioned reasons and in order to avoid ambiguity when assigning linguistic variables, a total of 1'816 students (out of 3'141) from the autonomous province of Bozen – South Tyrol are not part of our regression analyses, as they had missing information on both, the high school language and the communication language.

How assignment methods are distributed over languages is reported in Table 1: we could exploit the information on high schools for the assignment of the Italian language with 95.5% of the linguistic information stemming from the first method. Note also that the subsample of students with English as their mother tongue is very small, which is mostly due to the elimination of monolingual graduate programs and incoming exchange students from our original sample. Also, we have a limited number of bi- and trilingual students in our sample who reported highs school degrees in more than one native language.

In Table 2 we test whether the alternative language assignment methods are consistent and can confirm a strong correlation between the three proxies for a student's native language.

Language levels follow the *Common European Framework of Reference for Languages*

[1]This concept is however to be interpreted in a larger sense of "kindred" language.

[2]This also reflects the University policy to determine the student's first language (L1) at the moment of enrollment.

[3]The second method however was not suitable to determine the linguistic background of those born in the autonomous province of Bozen – South Tyrol, which account for 47.6% of our sample, as the around two thirds of the population speak German and less than one quarter speaks Italian as their first language.

Table 1: Distribution of assignment methods over languages

Native Language	**High School**		**Country of Origin**		**Communication**		**Total**	
	No.	Col %	No.	Col %	No.	Col %	No.	Col %
English	38	38.4	24	24.2	37	37.4	99	100.0
English and German	10	100.0	0	0.0	0	0.0	10	100.0
English and Italian	12	100.0	0	0.0	0	0.0	12	100.0
German	1,309	64.3	651	32.0	75	3.7	2,035	100.0
German and Italian	34	100.0	0	0.0	0	0.0	34	100.0
German, Italian and English	2	100.0	0	0.0	0	0.0	2	100.0
Italian	2,289	95.5	48	2.0	59	2.5	2,396	100.0
Total	3,694	80.5	723	15.8	171	3.7	4,588	100.0

Table 2: Correlation between Native Language Assignment Methods

	High School	Country of Origin	Communication Language
High School	1		
Country of Origin	0.900***	1	
Communication Language	0.767***	0.771***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

from A1 (Beginner) to C2 (Proficiency) and imported into the main dataset from a separate file containing all information on language exams and language certificates, including the date of upload. We decreased the date of upload by 90 days assuming that the language skills were already present during that time span before giving the exam[4].

Descriptives Not all of the 161'117 exams were "Passed", 29.15% of them feature status "Not passed", "Not presented", "Withdrawn", "To do" and "Other", which are not graded and therefore do not enter our regression analyses. Table 3 reports the overall faculty split with more than two thirds of the students enrolled in the Faculty of Economics and Management.

Table 3: Students by Faculty

Faculty	No.	%
Faculty of Computer science	381	5.8
Faculty of Design and Art	1,095	16.6
Faculty of Economics and Management	4,463	67.6
Faculty of Science and Technology	665	10.1
Total	6,604	100.0

The sex-ratio is quite balanced overall, with a slightly higher female presence in the

[4]Robustness analyses with alternative certificate dates do not impact the significance level of our results and are available upon request.

faculties of Design and Art and Economics and Management and a significantly lower female participation rate for the faculties of Computer Science and Science and Technology.

Table 4: Female participation rate by Faculty

	mean
Faculty of Computer science	0.165
Faculty of Design and Art	0.643
Faculty of Economics and Management	0.618
Faculty of Science and Technology	0.284
Total	0.563

Table 5 summarizes the average grades achieved by students born in the province of Bozen-Bolzano and those born elsewhere. A two-sample z-test shows that the difference of 0.6448 between entrants and locals is statistically significant (z=26.2966, p=0.0000) which means that students born elsewhere perform systematically better than their local colleagues.

Table 5: Selection bias local vs. non-local students

	No. students	Mean	SD
Not born in BZ	3479	24.900	2.28
Born in BZ	3187	24.255	2.17
Total	6666	24.592	2.25

The language level of the student at the point of taking an exam was set by matching the most recent language certificate date available from a secondary database with the date of the exam. Table 6 shows the language levels for passed exams given in Italian, German and English respectively[5].

Table 6: Language Level by Exam Language

Language Level	**Exam in English**		**Exam in German**		**Exam in Italian**		**Total**	
Language Level A1 or A2	3	0.0%	2	0.0%	1	0.0%	6	0.0%
Language Level B1 or B2	13,003	69.6%	2,298	18.3%	2,824	16.8%	18,125	37.8%
Language Level C1 or C2	4,826	25.8%	861	6.9%	1,932	11.5%	7,619	15.9%
Native Speaker	840	4.5%	9,371	74.8%	12,021	71.6%	22,232	46.3%
Total	18,672	100.0%	12,532	100.0%	16,778	100.0%	47,982	100.0%

[5]The minimum entry requirement for the third language is A0, which has to be brought up to B1 by the end of the first academic year. The small number of exams under A1 and A2 are rare cases of students who took exams with basic linguistic skills during their first year of studies.

Overall, by the end of their careers, non-native students tend to converge towards a medium-high level of proficiency in all three languages. Tables 7 show the maximum language level (L2 only) achieved by students at the end of their courses. For the sake of comparison, it is to be mentioned that the University automatically assigns the level C1 to the high school language considering it his or her native language. Thus, we created a new variable for which we assigned CEFRL levels only to L2 and recoded the native level as an additional category.

Table 7: Maximum Language Levels at graduation

	English		German		Italian	
Level	N	%	N	%	N	%
B1	24	2	42	6	18	3
B2	662	52	421	59	256	41
C1	548	43	247	35	340	55
C2	51	4	1	2	4	1
Total	1'285	100	711	100	618	100

Our dataset contains 6'610 unique exam clusters in terms of Exam Code, session year and Language of the Exam and a total of 94'033 exams have information on both, the exam language and the student's mother tongue. Quite in-line with the trilingual educational objectives of the University, 31% of the exams were given in the students' native language and 69% were given in a foreign language.

Professor types. In order to account for the instructor's characteristics, we created "Professor Types" which are grouped by gender, country of birth and chosen language of communication. On total we have 56 different professor-types, where one type shares the same gender, country of birth and language of communication. The most frequent types are shown in Table 1 (passed exams only) with the number of exams they graded on the y-axis. The two most frequent types are male professors who were born in Italy and chose Italian as their language of communication and male professors born in Germany with German as their preferred language of communication. On the third place we find female professors born in Italy with Italian as their preferred language.

Overall, female instructors are strikingly underrepresented in our sample as shown in Table 8, which reports the female instructor participation rate by faculty.

Grade Distribution. Figure 2 shows the distribution of exam marks for each session faculty. Most of the exams were taken with the Faculty of Economics and Management

Figure 1: Distribution Professor Types

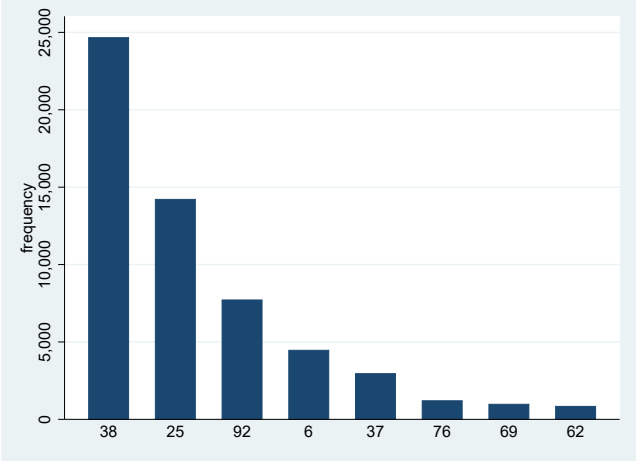


Table 8: Female Professor-Exams by Faculty

	mean
Faculty of Computer science	0.046
Faculty of Design and Art	0.235
Faculty of Economics and Management	0.212
Faculty of Science and Technology	0.154
Total	0.203

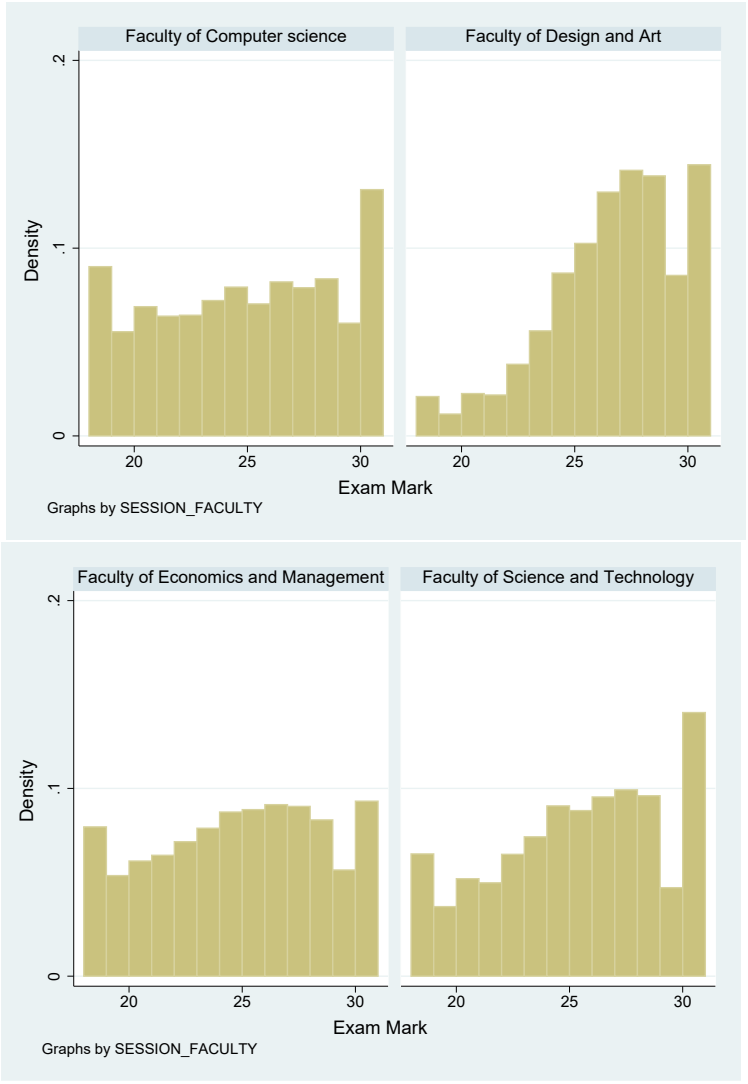
which exhibits a quite uniform distribution of grades, whereas the other three faculties have a tendency of assigning higher grades and/or award the highest grade with honors. This makes the case for the inclusion of additional session faculty fixed effects in our regression analyses.

6 Results

6.1 Exam Language Selection Bias - Cohort Analysis

First, we were interested in knowing whether the exam portfolio of German, English and Italian natives exhibit similar patterns. To this end we created a total of 165 *cohorts* by course and year of enrollment and compare students within the same cohort. 56 cohorts contain English native speakers, 141 cohorts are populated by German natives and in 151 cohorts we find Italian natives. Figure 3 compares the average exam language portfolio of Italian and German students, our main sample of interest. The University policy aims at a trilingual study plan where Italian, English and German as the exam language should be equally represented. However, we see that the portfolio composition of both Italian

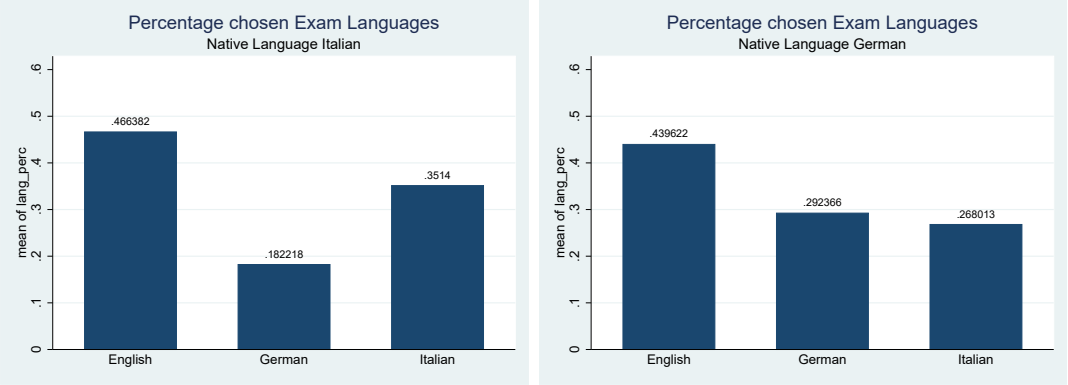
Figure 2: Grade Distribution over Session Faculties



and German natives tend towards the inclusion of exams in English for the main part and the own native language as the second choice. Furthermore, Italian natives have a significantly higher share of exams in L1 than their German colleagues.

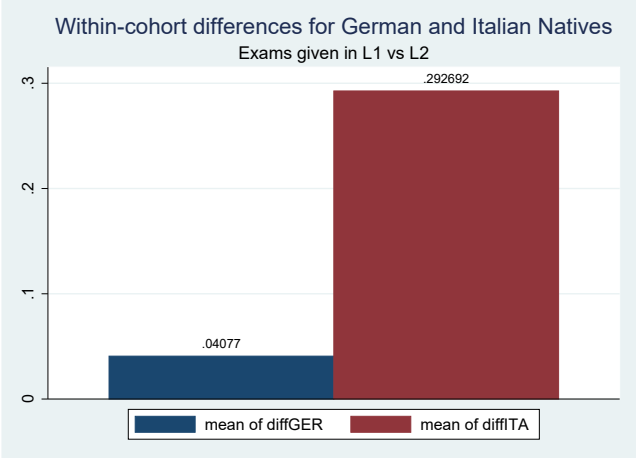
Figure 4 shows the average within-cohort differences between exams in L1 and exams in L2 for students with German versus Italian mother tongue. For this analysis we considered only the cohorts with both Italian and German natives enrolled. Under the hypothesis of no linguistic bias in the exam language portfolios compositions of the students, we should not observe any difference (i.e. no bars for both languages), as according to the educational mission of the University, L1 and L2 should account for one third regardless of the student's native language. However, we observe also in this descriptive representation

Figure 3: Exam Language Portfolio by Native Speakers



that even if both groups tend towards taking exams in their L1, Italian natives have a significantly higher proportion of exams in their L1, which is expressed by the higher average within-cohort difference between L1 and L2 exams for Italian students (red bar in 4).

Figure 4: Average within-cohort differences L1-L2



6.2 Regression Analyses

For an initial exploratory purpose we ran a pooled regression model (Table 9) with the exam grades as dependent variable and standard errors clustered at the individual student level. Grades in Italy span from 0 to 30 with honors, which was translated as a grade range from 0 to 31. To pass an exam, a minimum grade of 18 must be achieved, so we normalized marks on passed exams over the range of 18-31.

Female students appear to perform slightly better than their male colleagues in the

Table 9: Pooled Regression, SE clustered at the individual level

	(1) Exam Mark	(2) Exam Mark	(3) Exam Mark	(4) Exam Mark	(5) Exam Mark	(6) Exam Mark
Female student	0.0253***	0.00272	0.00274	0.00829	0.0106*	0.0398***
	(0.00534)	(0.00541)	(0.00483)	(0.00527)	(0.00521)	(0.00540)
Native English	-0.0668**	-0.0661*	-0.0694*	-0.0440	-0.0357	-0.0426*
	(0.0231)	(0.0323)	(0.0273)	(0.0276)	(0.0272)	(0.0198)
Native Italian	0.0298***	-0.00575	-0.0207**	-0.0215**	-0.00308	0.00873
	(0.00521)	(0.00692)	(0.00642)	(0.00714)	(0.00500)	(0.00773)
Exam in English	0.0690***	0.0706***	0.0915***	0.0504***	0.0509***	0.0455***
	(0.00429)	(0.00421)	(0.00363)	(0.00511)	(0.00512)	(0.00506)
Exam in Italian	0.0315***	0.0345***	0.0308***	-0.00797	-0.00764	-0.0114*
	(0.00327)	(0.00330)	(0.00311)	(0.00546)	(0.00546)	(0.00541)
Exam not in mother tongue	-0.0906***	-0.0939***	-0.0955***	-0.0890***	-0.0889***	-0.0850***
	(0.00359)	(0.00353)	(0.00326)	(0.00381)	(0.00382)	(0.00374)
External Exams	0.118***	0.100***				
	(0.00774)	(0.00798)				
Constant	0.516***	-0.150***	0.0431	-0.0522	0.144*	0.516***
	(0.00545)	(0.0239)	(0.0320)	(0.0616)	(0.0645)	(0.0287)
N	65324	61738	61334	41197	41197	43545
High School Grade	No	Yes	Yes	Yes	Yes	No
High School Type	No	Yes	Yes	Yes	No	Yes
Cohort	No	No	Yes	Yes	Yes	Yes
Session Faculty	No	No	Yes	No	No	No
Professor Type	No	No	No	Yes	Yes	Yes

Reference categories are Native German and Exam in German
Standard errors in parentheses
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

first model which does not include any additional controls. However, once we control for high school characteristics, this competitive gender advantage disappears. In model six, we only control for the high school type and not the grade which results again in a significant positive effect of female gender on university grades. If we assume that the high school grade is a proxy for the individual skill level, then the gender effect is absorbed by the higher skill set of female students and gender does not add additional effects during tertiary education. Next, the effects of being native English or native Italian as opposed to native German (baseline), does not show significant robust effects throughout the six model specifications and disappear completely once we control for high school grade, cohort and professor type. This shows that that there are no significant *ex ante* disadvantages between students of different linguistic backgrounds. With regards to the exam language, English exams are graded consistently higher than German exams,

whereas the significance for higher grades on exams in Italian disappears when we increase our set of controls including also professor types. What remains robust throughout all model specifications are the negative effects on grades for exams that are not in the student's mother tongue. Grades of non-native exams are approximately 8.5% lower (around 1.2 grade points), even after including a strict set of controls in terms of high school grade and type, cohort, session faculty and professor type. Finally, in the first two models we also take a glance into the set of exams taken with other universities (Erasmus, Free Mover Programs, etc.) and find that grades are around 10% higher than those of exams taken in Bolzano. External exams in models (3) to (6) are excluded by constructions, as we control for University-specific fixed effects.

In order to account for each student's individual academic path, we then ran a fixed-effects panel regression model in Table 10. Again we find that the effect of giving an exam in a foreign language decreases grades on average by 9.8%. A similar effect emerges from model (3) with which we show the effect of the language level. Having a language level of C1 or C2 increases grades by around 3.2% and being a native speaker adds on average an extra 10 percentage points to the exam mark.

Table 10: Fixed-Effects Panel Regression

	(1) Exam Mark	(2) Exam Mark	(3) Exam Mark
Exam not in mother tongue	-0.0670***	-0.0983***	
	(0.00290)	(0.00364)	
Exam in English		0.0537***	0.0568***
		(0.00492)	(0.00685)
Exam in Italian		-0.00982	-0.0217**
		(0.00527)	(0.00716)
Language Level C1 or C2			0.0319***
			(0.00540)
Native Speaker			0.102***
			(0.00504)
Constant	0.796***	0.784***	0.690***
	(0.0133)	(0.0135)	(0.0173)
Number of Exams	43098	43098	31474
Number of students	4090	4090	4000
Year of Study	Yes	Yes	Yes
Professor Type	Yes	Yes	Yes

Reference categories are Exam in German and Language Level A1 to B2 (Basic)
Standard errors in parentheses
*$p < 0.05$, *$p < 0.01$, *$p < 0.001$

Table 11 isolates the effects of L1 versus L2 for native Germans and native Italians. Model (1) includes the exams taken by native Germans in L1-German and L2-Italian, model (2) considers exams for native Germans in L1-German and L2-English. Similarly, models (3) and (4) analyze the outcomes for native Italians in their mother tongue and L2-German and L2-English respectively. By comparing models (1) and (3), we find that being a native speaker increases scores significantly for both Germans and Italians, whereas the effect for holders of language levels C1 and C2 is significant only for German natives.

Table 11: Fixed-Effects Panel Regression: Between-native comparison

	(1) Native GER No ENG exams	(2) Native GER No ITA exams	(3) Native ITA No ENG exams	(4) Native ITA No GER exams
Language Level C1 or C2	0.0730***	0.0320**	0.0296	0.0188*
	(0.0133)	(0.0123)	(0.0166)	(0.00910)
Native Speaker	0.133***	0.0375***	0.0744***	0.0117*
	(0.0174)	(0.00982)	(0.0149)	(0.00580)
Constant	0.544***	0.619***	0.652***	0.645***
	(0.0202)	(0.0128)	(0.0213)	(0.0250)
Number of Exams	9657	10734	11089	14863
Number of students	1687	1749	1966	2132
Professor Type	Yes	Yes	Yes	Yes
Year of Study	Yes	Yes	Yes	Yes

Reference category is Language Level A/B (A1 to B2). Unibz exams only.
The first model refers to exams in German or in Italian taken by native Germans, the second column refers to exams in German or in English taken by native Germans, similarly for columns 3 and 4.
Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Language Acquisition Tools. The fixed-effects panel regressions in Tables 12 and 13 looks into the efficacy of language improvement tools. The Language Center of the Free University of Bozen-Bolzano offers personalized learning paths (*Lernwege*) which are free of charge and aimed at improving the student's first and second foreign language. The entry requirements for study programs at the University foresee a minimum level of B2 for two of the three languages of instruction and the internal language courses should bridge the gap for the third language with the final objective of reaching B2 for all three languages. Alongside internal language exams, students also bring different types of external language certifications as proof of their skills. These certifications must not be older than five years and are acquired either before enrollment or during the course of studies. The third, more informal language acquisition tool in our analysis is given by periods spent abroad as part of the study program. Similarly to how we proceeded with the assignment of language certifications, by considering the starting and end date of the exchange program, we were able to identify those students who already completed a period abroad in the linguistic area of the exam language.

Overall and irrespective of the type of language acquisition tool, having a language level of C1 or C2 in L2-Italian has a significant impact on grades for German natives (models (1) and (2)), whereas we do not find significant effects in terms of grades for native Italians and L2-German compared to the baseline level A1 to B2 (see Table 13). It is to be mentioned here that the Free University of Bozen-Bolzano requires the instructors not to penalize students based on their language skills, but to grade only the subject-specific contents of the exam. As a consequence the difference in L2 improvements should be due to objective improvements in terms of content apprehension thanks to higher linguistic skills in L2.

Table 12: Efficacy of Language Improvement Tools: German natives - exams in L2 (Italian)

	(1)	(2)	(3)
Language Level C1 or C2	0.0421*	0.0415*	
	(0.0182)	(0.0197)	
External Certificate		0.00185	
		(0.0236)	
Previous period abroad			0.0871***
			(0.0243)
Constant	0.457***	0.456***	0.464***
	(0.00738)	(0.0102)	(0.00115)
Number of Exams	4483	4483	8587
Number of students	997	997	1638

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: Efficacy of Language Improvement Tools: Italian natives - exams in L2 (German)

	(1)	(2)	(3)
Language Level C1 or C2	0.0506	0.0795**	
	(0.0277)	(0.0300)	
External Certificate		-0.0981**	
		(0.0335)	
Previous period abroad			0.102***
			(0.0184)
Constant	0.447***	0.488***	0.455***
	(0.00745)	(0.0161)	(0.00230)
Number of Exams	3023	3023	5659
Number of students	1141	1141	1625

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In models (2) of tables 12 and 13 we control for the origin of linguistic skill improvement and find no significant effects for German natives relatively to external language certificates and courses offered by the University. For Italian students on the other hand, it appears that external certificates are less effective in terms of grade improvement with respect to internal courses of German. The linguistic knowledge gained from exchange programs increases the average grade by approximately 8.7% for German natives and 10.2% for Italian natives (models (3)). Consider here that even though we included national exchange programs, these are much less frequent that exchange programs such as for example the Erasmus program. For a German native, having spent a period classified as "abroad" in an Italian area is, in fact, a national exchange with another Italian university.

Stratified Regression. Finally, we ran a stratified panel regression by study year for the subsample of Bachelor students. The non-native disadvantage shows a non-linear trend with a negative effect in the first study year, which is significant only at the 10% level, no significant at all in the second year and a strongly significant negative impact in the third study year. Considering what we described an analyzed so far, these results reflect an interesting narrative. The reason behind these findings, which might appear irregular at first sight, most likely lie in the numerous L2-elusion possibilities available to the students

during the first two study years. Namely, they could delay and anticipate exams based on the exam language, minimizing in this way the negative impact on their GPA. In the second year, participation in mobility programs is quite common, wherefore students are exempt from the trilingual obligation. In the third year, typically, the students have to deal with the remainder of the exams they could not delay, anticipate or take abroad, wherefore, not having a "way out", the non-native disadvantage is the strongest.

Table 14: Stratified Panel Regression - Bachelor program

	(1) 1st year	(2) 2nd year	(3) 3rd year
Exam not in mother tongue	-0.164*	-0.123	-0.124***
	(0.0715)	(0.0734)	(0.0367)
Exam in English	0.0899***	0.0391**	0.0195
	(0.0168)	(0.0141)	(0.0137)
Exam in Italian	0.0447*	-0.0486**	-0.100***
	(0.0194)	(0.0148)	(0.0131)
Constant	0.726***	0.742***	0.728***
	(0.0735)	(0.0756)	(0.0435)
Number of Exams	8294	9825	8573
Number of students	3238	2945	2392
Professor Type	Yes	Yes	Yes
Language Level	Yes	Yes	Yes

Reference categories are Exams in German
Standard errors in parentheses
*$p < 0.05$, *$p < 0.01$, *$p <$ 0.001

Failed attempts Another interesting area of study for the effect of native language on performance is the number of failed attempts to give a certain exam. To this end we created a new variable which reports the number of failed attempts at the exam code and exam language level. Recall that the exam language is always the same throughout the whole session year. Also, in 2013 the University implemented a reform allowing students

to register for an exams three times instead of twice throughout the academic year[6]. Failed attempts include failed exams, students that started to write the exam, but decided to withdraw from it and unjustified absences at the day of the exam. In table 15 we ran three Poisson regressions with random effects which allows us to control for individual latent determinants at the student level without excluding singleton panel observations, i.e. only one failed attempt per exam code and language. In model (1), the Incidence Rate Ratio (IRR) for exams given in a language different to the native language is greater than one and strongly significant, meaning that exams in a second language compared to exams in the mother tongue, while holding the other variables constant in the model, are expected to have a rate 1.057 greater for failed attempts. Somewhat surprisingly, failed attempts slightly decrease after the 2013 reform and the effect is robust throughout all model specifications. Looking further into detail, we find that Italians appear to have a slightly higher tendency for failed attempts with respect to German natives, but when we then look into the interaction term with the exams not in mother tongue, we find that for exams in L2, Italian natives fail less often that their German colleagues. Part of the effect might also be due to an overall higher failure rate for exams in Italian, which is shown in model (3).

[6]As a consequence of this reform, also teaching contracts were extended in order to include a full academic year. Whereas before 2013 teaching contracts for second-semester courses ended in September, after the reform they run up to the end of February. First-semester courses run from the beginning of October to the end of September of the subsequent calendar year.

Table 15: Failed attempts - Poisson regression (Incidence Rate Ratio)

	(1)	(2)	(3)
Exam in L2	1.057***	1.152***	
	(0.02)	(0.03)	
Post 2013 Reform	0.954***	0.961**	0.959**
	(0.01)	(0.01)	(0.01)
Native English		1.416***	1.161***
		(0.07)	(0.05)
Native Italian		1.070*	0.969*
		(0.03)	(0.01)
Exam in L2 × Native English		0.688***	
		(0.06)	
Exam in L2 × Native Italian		0.889***	
		(0.03)	
Exam in English			1.087***
			(0.02)
Exam in Italian			1.071***
			(0.02)

Exponentiated coefficients; Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6.2.1 Elusion Strategies for L2

Internal elusion strategies. A voluntary mechanism to deviate from the student language portfolio prescribed by the University, could consist in postponing the exam until a later date, hoping for a change in the teaching language in the subsequent academic year. The decision to take an exam at a later date with respect to the colleagues of the cohort could also depend on the need of additional time to improve the necessary linguistic skills to be ready for the exam. Even though students should be alien to any details on how teaching resources are managed by the University, it could well be that they are able to retrieve insider information on future changes of instructors or probabilities on such changes. Also, student could decide to anticipate the exam in L2 if in the current academic year the exam is offered in a language preferred by the student (native or the other L2). To

analyze potential shirking behavior, we considered again the cohort level and determined the first possible session year and the corresponding exam language for that session. We then classified the exams as "switching", when the exam language changed from one academic year to the next. In Table 16 we ran three probit models with the likelihood of delaying the exam to a later academic year as the dependent variable. Postponers are those students who took the exam one or more academic years later that their cohort peers. To distinguish this phenomenon of internal shirking from other typologies of linguistic elusion strategies, we consider only the exams taken with the University. Also, we include only cohorts with more than 70 students. Within our subsample we find a general tendency towards delaying those exams that are taught in L2, for example to gain some time to improve linguistic skills. The interaction term between "Exam not MT" and the exam being of "switcher" type is also positive and significant. Together with the non-significant coefficient for the switching type only, the results suggest a strategic elusive behavior in terms of choosing the exam language for both German and Italians alike.

Table 16: Likelihood of delaying an exam - Basic Probit Models

	(1) Full Sample	(2) Native German	(3) Native Italian
Exam not MT=1	0.265***	0.331***	0.218***
	(0.0228)	(0.0366)	(0.0277)
Language Switch=1	0.0578	0.0824	0.0442
	(0.0495)	(0.0680)	(0.0738)
Exam not MT=1*Language Switch=1	0.275***	0.270***	0.271***
	(0.0528)	(0.0745)	(0.0762)
Constant	-1.028***	-1.085***	-0.993***
	(0.0241)	(0.0355)	(0.0326)
Observations	28273	13311	14359

Standard errors in parentheses
Standard errors clustered at the individual student level
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As a second mechanism of internal shirking, students could also manage to strategically anticipate the exam. To this end we ran a similar analysis in 17 and found that exams in L2 have a lower tendency to be taken early than the study plan foresees which is again

perfectly in line with the narrative of a time-intensive language acquisition process. Again, we observe a positive and significant interaction for exams in L2 which are of "switching" type indicating that speculation on or insider information about language changes in the future for a certain exam is part of the students' decision model regarding the exam date. We also find an overall tendency to avoid switching for exams that change the exam language from one academic year to the other. We assume that anticipation, as opposed to postponing an exam, requires a higher level of decision-making effort and also a higher level of risk. If the exam is known to change the exam language in the subsequent year it also means that the instructor changes. Hence, risk aversion could lead students to avoid this kind of decision-making altogether if it does not involve language issues.

Table 17: Likelihood of exam anticipation - Basic Probit Models

	(1) Full Sample	(2) Native German	(3) Native Italian
Exam not MT=1	-0.354***	-0.331***	-0.383***
	(0.0285)	(0.0432)	(0.0375)
Language Switch=1	-0.352***	-0.389***	-0.323**
	(0.0593)	(0.0776)	(0.0984)
Exam not MT=1*Language Switch=1	0.576***	0.625***	0.534***
	(0.0667)	(0.0937)	(0.101)
Constant	-1.243***	-1.219***	-1.279***
	(0.0257)	(0.0410)	(0.0341)
Observations	23902	11223	12139

Standard errors in parentheses
Standard errors clustered at the individual student level
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Abroad elusion strategies. The second mechanism of avoidance for L2 exams with the University is taking the exam abroad or with another Italian university. Table 18 shows the likelihood of taking an exam elsewhere if the first available exam language at the University was not the student's mother tongue. Also here, we observe a consistent positive effect for the full sample and also the three linguistic subgroups with the marginal percentage effects reported in table 19.

Table 18: Likelihood of taking the exam abroad - Basic Probit Models

	(1) Full Sample	(2) Native German	(3) Native Italian
Exam not MT=1	0.337***	0.252***	0.424***
	(0.0323)	(0.0631)	(0.0365)
Constant	-2.200***	-2.294***	-2.132***
	(0.0345)	(0.0640)	(0.0413)
Observations	43407	21236	21036

Standard errors in parentheses
Standard errors clustered at the individual student level
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 19: Probability of taking the exam abroad - Marginal Effects

	(1) Full Sample	(2) Native German	(3) Native Italian
Exam not MT (1 vs 0)	0.0173	0.00969	0.0273
	(0.00159)	(0.00216)	(0.00247)
Observations			

Marginal effects; Standard errors in parentheses
(d) for discrete change of dummy variable from 0 to 1

falegnameriacrema;libero.it

6.3 Conclusions

Our findings presented here indicate that there are no significant asymmetries ex ante in terms of grades between Italian and German natives. However, we do observe an average loss of more than one grade point out of 30 for exams that are not taken in the student's mother tongue. With these results in mind, we conducted a series of fixed-effect panel regressions on the full sample and on subgroups in order to get a better idea of the drivers and interaction effects of linguistic asymmetries in academic performance. Somewhat unsurprisingly, a higher language level (language certificates from A1 to C2 as per the

CEFRL) held at the point of taking an exam which is not in the student's mother tongue has a positive impact on grades. Furthermore, our results show that students seem to particularly benefit from periods abroad completed prior to an exam taken in L2. Periods abroad also increase the overall GPA of the students as they tend to be higher graded than local exams, which leads to a twofold advantage for students participating in mobility programs. We also analyze elusion behavior with respect to the trilingual language policy of the Free University of Bozen-Bolzano and find that students strategically postpone and anticipate exams based on the exam language. Also, they have a higher tendency of redirecting exams which would have been offered in their second language towards mobility programs abroad. In conclusion, our results underline the importance for governing bodies of higher education institutions to explicitly address the topic of non-native learning and develop targeted policies in order to avoid inequalities in terms of grades between students of different linguistic backgrounds. Furthermore, in terms of post-graduate effects, job application procedures should put particular attention on integrating a standardized GPA interpretation with further information on the applicant's linguistic curriculum during the years of education in order to avoid excluding those candidates with a lower GPA solely due to an educational path undertaken in a non-native language.

A Robustness Analyses - Excluding Pre-2011 enrollment

Table 20: Pooled Regression, SE clustered at the individual level, Post-2011 enrollment only

	(1) Exam Mark	(2) Exam Mark	(3) Exam Mark	(4) Exam Mark	(5) Exam Mark	(6) Exam Mark
Female student	0.0206**	-0.000168	0.00212	0.00674	0.00888	0.0400***
	(0.00680)	(0.00667)	(0.00666)	(0.00721)	(0.00711)	(0.00727)
Native English	-0.0289	-0.0385	-0.0289	0.00605	-0.0160	-0.0111
	(0.0236)	(0.0575)	(0.0592)	(0.0610)	(0.0574)	(0.0257)
Native Italian	0.0291***	-0.0161*	-0.0218**	-0.0204*	0.00581	0.0105
	(0.00660)	(0.00794)	(0.00749)	(0.00832)	(0.00686)	(0.00899)
Exam in English	0.0917***	0.0993***	0.103***	0.0401***	0.0400***	0.0328***
	(0.00499)	(0.00507)	(0.00491)	(0.00769)	(0.00770)	(0.00741)
Exam in Italian	0.0386***	0.0464***	0.0417***	-0.0339***	-0.0330***	-0.0393***
	(0.00441)	(0.00455)	(0.00443)	(0.00817)	(0.00817)	(0.00787)
Exam not in mother tongue	-0.0890***	-0.0918***	-0.0912***	-0.0828***	-0.0818***	-0.0798***
	(0.00438)	(0.00455)	(0.00442)	(0.00515)	(0.00517)	(0.00489)
External Exams	0.109***	0.0904***				
	(0.00875)	(0.00915)				
Constant	0.508***	-0.239***	0.00204	0.107	0.332***	0.833***
	(0.00724)	(0.0291)	(0.0578)	(0.0578)	(0.0485)	(0.0408)
Observations	34985	31904	31904	20930	20930	23049
High School Grade	No	Yes	Yes	Yes	Yes	No
High School Type	No	Yes	Yes	Yes	No	Yes
Cohort	No	No	Yes	Yes	Yes	Yes
Session Faculty	No	No	Yes	No	No	No
Professor Type	No	No	No	Yes	Yes	Yes

Table 21: Fixed-Effects Panel Regression - Post-2011 enrollment only

	(1) Exam Mark	(2) Exam Mark	(3) Exam Mark
Exam not in mother tongue	-0.0509***	-0.0899***	
	(0.00368)	(0.00479)	
Exam in English		0.0436***	0.0457***
		(0.00737)	(0.00742)
Exam in Italian		-0.0338***	-0.0343***
		(0.00768)	(0.00770)
Language Level C1 or C2			0.0287***
			(0.00581)
Native Speaker			0.0995***
			(0.00523)
Constant	0.849***	0.841***	0.733***
	(0.0349)	(0.0341)	(0.0333)
Number of Exams	22711	22711	22599
Number of students	2756	2756	2745
Year of Study	Yes	Yes	Yes
Professor and TA Type	Yes	Yes	Yes

Reference categories are Exam in German and Language Level A1 to B2 (Basic)

Table 22: Fixed-Effects Panel Regression: Between-native comparison - Post-2011 enrollment only

	(1) Native GER No ENG exams	(2) Native GER No ITA exams	(3) Native ITA No ENG exams	(4) Native ITA No GER exams
Language Level C1 or C2	0.0446**	0.0197	0.0175	0.0232*
	(0.0143)	(0.0135)	(0.0172)	(0.00941)
Native Speaker	0.149***	0.0538***	0.00134	0.0115
	(0.0211)	(0.0109)	(0.0183)	(0.00614)
Constant	0.613***	0.652***	0.648***	0.652***
	(0.0294)	(0.0180)	(0.0220)	(0.0253)
Number of Exams	5586	6610	7527	11004
Number of students	1005	1061	1423	1594
Professor Type	Yes	Yes	Yes	Yes
Year of Study	Yes	Yes	Yes	Yes

Reference category is Language Level A/B (A1 to B2). Unibz exams only.
The first model refers to exams in German or in Italian taken by native Germans, the second column refers to exams in German or in English taken by native Germans, similarly for columns 3 and 4.
Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 23: Efficacy of Language Improvement Tools: German natives - exams in L2 (Italian) - Post-2011 enrollment only

	(1)	(2)	(3)
Language Level C1 or C2	0.0349	0.0369	
	(0.0205)	(0.0221)	
External Certificate		-0.00688	
		(0.0262)	
Previous period abroad			0.0837*
			(0.0372)
Constant	0.456***	0.458***	0.464***
	(0.00814)	(0.0118)	(0.00224)
Number of Exams	3701	3701	3792
Number of students	861	861	888

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 24: Efficacy of Language Improvement Tools: Italian natives - exams in L2 (German)

	(1)	(2)	(3)
Language Level C1 or C2	0.0361	0.0665*	
	(0.0286)	(0.0314)	
External Certificate		-0.0993**	
		(0.0350)	
Previous period abroad			0.108***
			(0.0251)
Constant	0.452***	0.494***	0.444***
	(0.00741)	(0.0166)	(0.00431)
Number of Exams	2876	2876	2918
Number of students	1083	1083	1110

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

References

Abutalebi, J. and Clahsen, H. (2017). Memory retrieval and sentence processing: Differences between native and non-native speakers. *Bilingualism: Language and Cognition*, 20(04):657–658.

Akerlof, G. A. and Kranton, R. E. (2003). Identity and Schooling: Some Lessons for the Economics of Education. *Journal of Economic Literature*, 40(4):1167–1201.

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. Number January. NY: Columbia University Press.

Bookheimer, S. (2002). Functional MRI of Language: New Approaches to Understanding the Cortical Organization of Semantic Processing. *Annual Review of Neuroscience*, 25(1):151–188.

Brügger, B., Lalive, R., and Zweimüller, J. (2009). Does Culture Affect Unemployment? Evidence from the Röstigraben. *SSRN eLibrary*.

Cappellari, L. and Di Paolo, A. (2018). Bilingual schooling and earnings: Evidence from a language-in-education reform. *Economics of Education Review*, 64(9431):90–101.

Cappellari, L., Lucifora, C., and Pozzoli, D. (2012). Determinants of grades in maths for students in economics. *Education Economics*, 20(1):1–17.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9):2633–2679.

Clahsen, H. and Felser, C. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences*, 10(12):564–570.

Coleman, J. S. (1966). Equality of Educational Opportunity. Technical report, U.S. Dept. of Health, Education, and Welfare, Office of Education.

Costa, A., Vives, M., and Corey, J. D. (2017). On Language Processing Shaping Decision Making. *Current Directions in Psychological Science*, 26(2):146–151.

Cunnings, I. (2017). Parsing and Working Memory in Bilingual Sentence Processing. *Bilingualism: Language and Cognition*, 20(4):659–678.

Danzer, A. M., Feuerbaum, C., Piopiunik, M., and Woessmann, L. (2018). Growing Up in Ethnic Enclaves: Language Proficiency and Educational Attainment of Immigrant

Children. *IZA Discussion Paper*, (11608).

De Paola, M. (2009). Does teacher quality affect student performance? Evidence from an Italian university. *Bulletin of Economic Research*, 61(4):353–377.

Gazzola, M., Grin, F., and Wickström, B.-A. (2015). A Concise Bibliography of Language Economics. *CESIFO Working Paper*, (5530):1–61.

Grabner, R. H., Saalbach, H., and Eckstein, D. (2012). Language-Switching Costs in Bilingual Mathematics Learning. *Mind, Brain, and Education*, 6(3):147–155.

Griffith, A. L. and Rask, K. N. (2014). Peer effects in higher education: A look at heterogeneous impacts. *Economics of Education Review*, 39:65–77.

Hanushek, E. A. (1979). Conceptual and Empirical Issues in the Estimation of Educational Production Functions. *The Journal of Human Resources*, 14(3):351.

Hanushek, E. A. (1995). Interpreting Recent Research on Schooling in Developing Countries. *The World Bank Research Observer*, 10(2):227–246.

Hoffmann, F. and Oreopoulos, P. (2009). Professor qualities and student achievement. *Review of Economics and Statistics*, 91(1):83–92.

Hogan-Brun, G. and Polezzi, L. (2018). Mother tongue matters: growing up in a multilingual world.

Isphording, I. E. (2015). Language and Labor Market Success. *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*, pages 260–265.

Leiner, H. C., Leiner, A. L., and Dow, R. S. (1991). The human cerebro-cerebellar system: its computing, cognitive, and language skills. *Behavioural Brain Research*, 44(2):113–128.

Mehler, A., Zlatkin-Troitschanskaia, O., Hemati, W., Molerov, D., Lücking, A., and Schmidt, S. (2018). Integrating Computational Linguistic Analysis of Multilingual Learning Data and Educational Measurement Approaches to Explore Learning in Higher Education. In *Positive Learning in the Age of Information*, pages 145–193. Springer Fachmedien Wiesbaden, Wiesbaden.

Oosterbeek, H. and Van Ewijk, R. (2014). Gender peer effects in university: Evidence from a randomized experiment. *Economics of Education Review*, 38:51–63.

Porchea, S. F., Allen, J., Robbins, S., and Phelps, R. P. (2017). Predictors of Long-Term Enrollment and Degree Outcomes for Community College Students: Integrating Academic, Psychosocial, Socio-demographic, and Situational Factors. *The Journal of Higher Education*, 81(6):680–708.

Regester, D. and Norton, M. K. (2018). The Salzburg Statement for a Multilingual World.

European Journal of Language Policy, 10(1):156–162.

Robbins, S. B., Allen, J., Casillas, A., Peterson, C. H., and Le, H. (2006). Unraveling the differential effects of motivational and skills, social, and self-management measures from traditional predictors of college outcomes. *Journal of Educational Psychology*, 98(3):598–616.

Saiz, A. and Zoido, E. (2005). Listening to what the world says: Bilingualism and earnings in the United States. *Review of Economics and Statistics*, 87(3):523–538.

Schultz, T. W. (1960). Capital Formation by Education. *Journal of Political Economy*, 68(6):571–583.

Sutter, M., Angerer, S., Glätzle-Rützler, D., and Lergetporer, P. (2015). The Effect of Language on Economic Behavior: Experimental Evidence from Children's Intertemporal Choices. *IZA Discussion Paper*, (9383).

Thiele, T., Singleton, A., Pope, D., and Stanistreet, D. (2016). Predicting students' academic performance based on school and socio-demographic characteristics. *Studies in Higher Education*, 41(8):1424–1446.

Tinto, V. (1997). Classrooms as Communities: Exploring the Educational Character of Student Persistence. *The Journal of Higher Education*, 68(6):599.

Todd, P. E. and Wolpin, K. I. (2003). On the Specification and Estimation of the Production Function for Cognitive Achievement. *The Economic Journal*, 113(485):F3–F33.

Venkatraman, V., Soon, C. S., Chee, M. W., and Ansari, D. (2006). Effect of language switching on arithmetic: A bilingual fMRI study. *Journal of Cognitive Neuroscience*, 18(1):64–74.