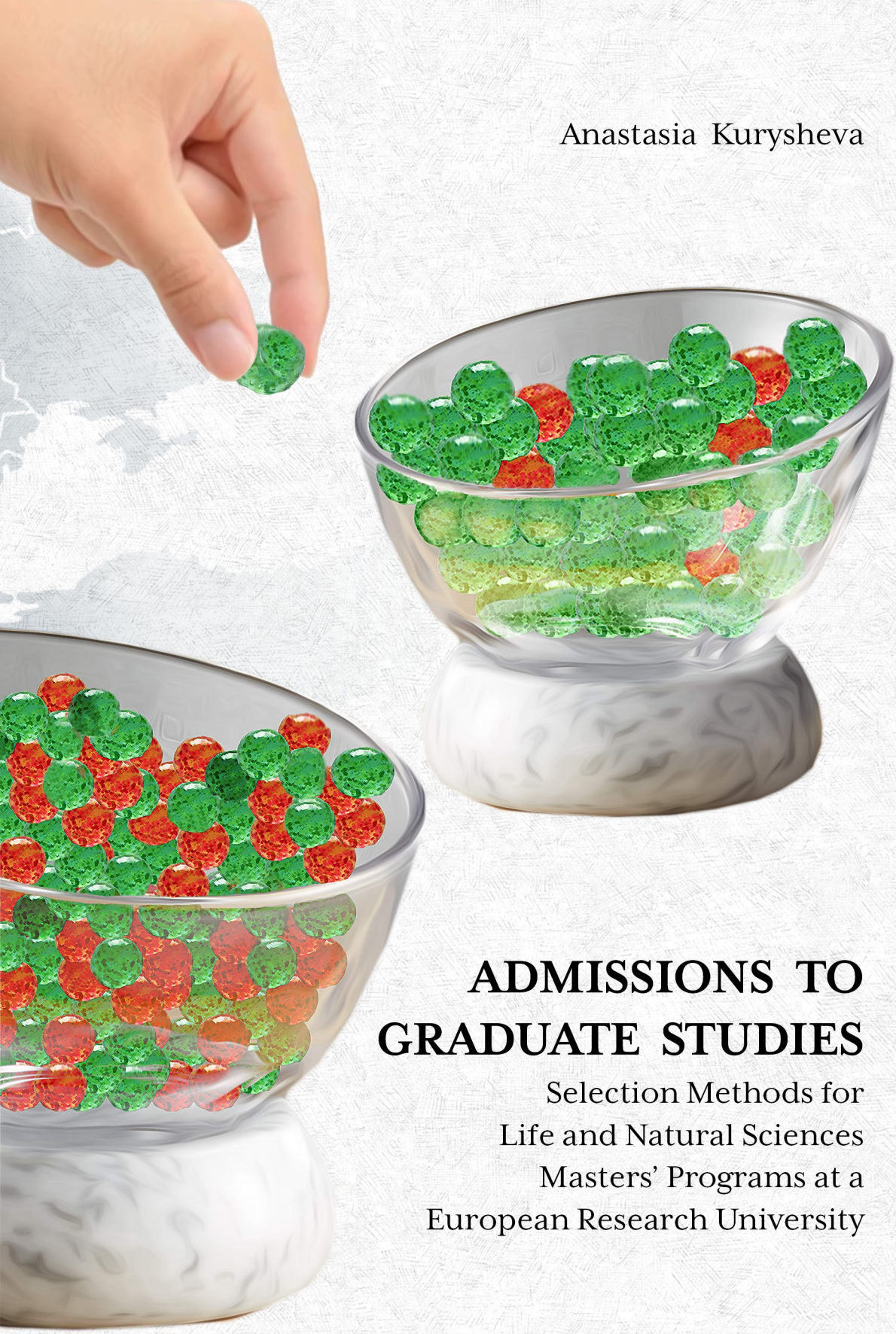
Anastasia Kurysheva



ADMISSIONS TO GRADUATE STUDIES

Selection Methods for Life and Natural Sciences Masters' Programs at a European Research University

Admissions to Graduate Studies:

Selection Methods for Life and Natural Sciences Masters' Programs at a European Research University

Anastasia Kurysheva

Admissions to Graduate Studies:
Selection Methods for Life and Natural Sciences Masters' Programs at a European Research University

ISBN: 978-94-6458-681-7
DOI: https://doi.org/10.33540/1501

Author: Anastasia Kurysheva
Cover design and layout: Iliana Boshoven-Gkini | www.AgileColor.com
Printed by: Ridderprint | www.Ridderprint.nl

Admissions to Graduate Studies: Selection Methods for Life and Natural Sciences Masters' Programs at a European Research University

Toelating tot Graduate Studies: Selectiemethoden voor Masterprogramma's in de Levens - en Natuurwetenschappen bij een Europese onderzoeksuniversiteit

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

dinsdag 15 november 2022 des middags te 4.15 uur

door

Anastasia Iuryevna Kurysheva

geboren op 31 augustus 1988
te Tashkent, Oez. SSR

Promotoren:

Prof. dr. H.V.M. van Rijen

Prof. dr. M.C. van der Wende

Copromotor:

Dr. G. Dilaver

Beoordelingscommissie:

Prof. dr. J.A.G. van Strijp (voorzitter)

Prof. dr. T.J. ten Cate

Prof. dr. M.F. van der Schaaf

Prof. dr. L.D.N.V. Wijngaards

Prof. dr. G. Croiset

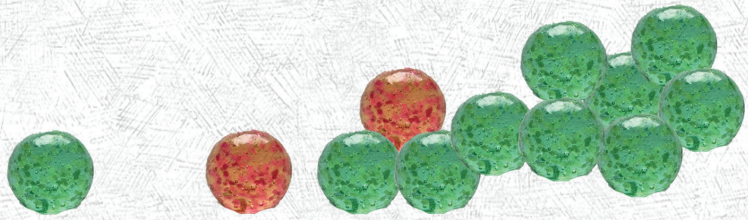
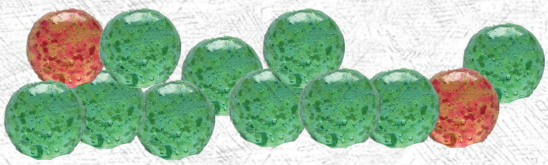
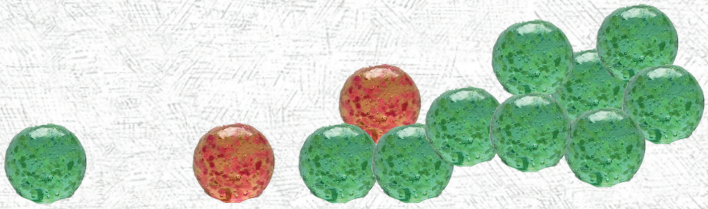


Table of Contents

1



General Introduction



1.1 Setting the Scene

Meet Eva, Jan, and Olivia. All three are about to finish their bachelors' studies. Eva is about to graduate from a Dutch research university, Jan finishes his bachelor's program at a Dutch university of applied sciences, and Olivia will soon obtain her bachelor's degree from a Spanish university. They wish to pursue a graduate degree in one of the Science, Technology, Engineering, and Math (STEM) fields, and all three chose to apply to the same selective master's research program at a major Dutch research university because of the program's strong research focus, internationally acknowledged reputation, and appealing career perspectives. After having sent their applications and while waiting for the letter from the program's admissions committee, the students hold their breath: They know that there are multiple applicants per place, and the chances of being rejected from this attractive master's program are high.

At the same time, the admissions committee is reviewing 120 applications they received for the 30 places on the program. Most of the applicants possess a relevant bachelor's degree and thus fulfil the minimum requirements for studying on this program. After having undergone the intensive rounds of evaluating, discussing, and having finally ranked all 120 applications based on the program's various selection criteria, the admissions committee was able to fill 29 of the 30 places. Tied for the last place in their ranking are Eva, Jan, and Olivia. To fill this remaining place, one more round of evaluation must be conducted. This round will hopefully highlight subtle but sensible differences between these three applicants and will indicate who out of the three has a better chance of becoming a successful student on their master's program.

The application files as well as the notes and memories from the admissions interviews provide the committee members with a wealth of information about the applicants. Cognizant of the weight of their decision, they want to make a well-informed and justifiable choice in selecting one candidate over the others. However, how do the committee members know which selection methods to use to make a distinction between these three candidates? Will their decision be based on scientific evidence and therefore justifiable? Will it be transparent to applicants?

In this thesis, I attempt to provide empirical insights aimed at assisting admissions committees in making these by-nature complex decisions.

1.2 Structure of the Chapter 1

Before delving into the design of this thesis, I will first describe the Dutch Higher Education context—in which the empirical studies were conducted—to assist the reader in considering which aspects of this thesis could be generalized to other national higher education systems. The description of the context includes several aspects. First, the background of the Dutch higher education system's trajectory toward a selective

admissions model in the second half of the 20th century is outlined. Continued by the consequences that the Bologna Process and the increasing internationalization had in this respect. Finally, these are put into the context of equity considerations which have become more important in light of a diversifying student population.

After presenting the context, the scope of the thesis is outlined. It begins by defining the broad theoretical and methodological perspectives on student selection undertaken in this thesis. Next, the research problem is addressed along with the theoretical and practical importance of conducting this research. Graduate study success is also defined within the framework of this thesis. Finally, the aims, research questions, and structure of the thesis as a whole are presented.

1.3 The Dutch Higher Education Context

Selective admissions decisions are challenging by nature, as they profoundly impact an individual's education, careers, and quality of life (Zwick, 2019). Consequently, evidence-based student selection decision making for objective, transparent, and fair selective admissions has been a prominent topic of interest for several decades. Some of these discussions emerged in the US already at the beginning of the 20th century (Karabel, 2006), but more commonly and also in Europe, these considerations came as a consequence of the massification of higher education in the second half of the 20th century. For example, in the Netherlands, the debate emerged in the 1970s, but it was not until 1999 that selective admissions started being implemented in practice. Currently, up to 10% of bachelors' programs and at least 40% of masters' programs apply selective admissions (van der Wende, 2020). This indicates that selective admissions are becoming more common at the master's level. To better understand why selective admissions are important as a societal topic in the Netherlands, an overview of the Dutch higher education system and the factors surrounding the admissions debate are presented below.

1.3.1 The Dutch Higher Education System

The Dutch higher education system is binary and predominantly made up of public institutions with only a few small specialized private institutions. The binary divide (see Figure 1) manifests in two major sectors of higher education: universities of applied sciences ("hogescholen", in Dutch) and research universities (including some university colleges, which offer selective international liberal arts & sciences bachelor programs; van der Wende, 2020). The universities of applied sciences offer higher professional education and have a limited, practice-oriented research portfolio (van der Wende, 2020). They grant professional degrees at associate, bachelors', and masters' levels. Research universities have large research portfolios and grant degrees from bachelor up to PhD level.

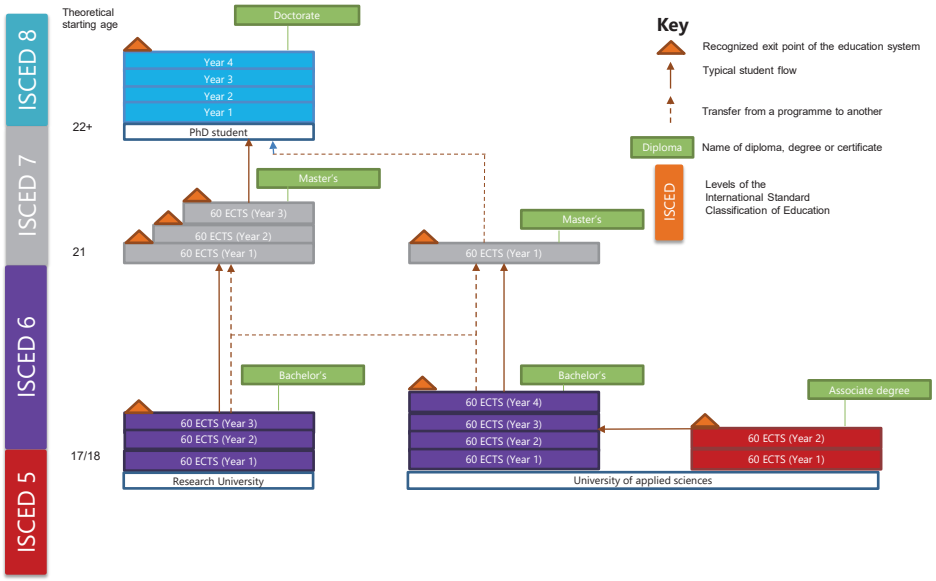


Figure 1 | *The Educational Pathways Leading to Masters' Degrees in the Dutch Higher Education System.*

Note. The diagram is adapted from *Diagram of education system: Netherlands*, by Education GPS, Organization for Economic Co-operation and Development (OECD), 2022 (https://gpseducation.oecd.org/CountryProfile?primaryCountry=NLD). Copyright 2020 by Education GPS.

As depicted in Figure 1, masters' programs in the Netherlands typically last one year (60 ECTS[1] points) in the social sciences and humanities, two years (120 ECTS points) in the life and natural sciences, and three years (180 ECTS points) in medical studies. A specific category of masters' programs—research masters' programs—was launched in 2003. These programs prepare students for a research career within and outside academia (Accreditation Organization of the Netherlands and Flanders [NVAO], 2016). Three main features that distinguish research masters' from taught (or academic) masters' are: (1) their positioning toward research-intensive education (e.g., research internships typically constitute components of such masters' programs), (2) duration (two years and 120 ECTS credits instead of one year and 60 ECTS credits), and (3) selectiveness (Snijder, 2016). In this thesis, the empirical studies are conducted on students in research masters' programs.

1.3.2 Pathways to Graduate Education in the Netherlands

The standard paths to a Dutch master's program are through a three-year bachelor's program from a Dutch research university (or university college as a department of a research university), a four-year bachelor's degree from a university of applied sciences, or the equivalent for both types of bachelors' degrees from foreign Higher Education Institution (HEI; Figure 1).

1 European Credit Transfer and Accumulation System

1.4 Arduous Journey toward Selective Admissions in the Netherlands

1.4.1 Previous Admissions Models: Open Admissions and Weighted Lottery

Previously, access to most Dutch higher education programs was granted based on an open admissions model, which means that any student with a qualifying secondary-level certificate may enroll into higher education (Williams & Wendler, 2020). From 1975 to 2001, that is prior to the introduction of the Bologna process (see next section), a limited number of programs in professional education (where the number of applicants exceeded the number of available places) used student selection, and universities applied the weighted lottery system in which chances of admission increased according to the student's secondary school average grade (Schripsema, 2017). The latter system was criticized, especially in 1996, after an academically outstanding student was repeatedly denied entry to a medical school (Goudappel, 1999). This resulted in changes in the system in 1999 when all applicants with a grade corresponding to a certain threshold (namely, an average for the final exam of 8 or above on the Dutch grading scale from 1 to 10) were admitted directly. The rest of the applicants were generally admitted through weighted lottery which was used to fill at least half of the available places left after admitting those through direct access (Goudappel, 1999; Kaiser & Vossensteyn, 2005, 2009). For the remaining places (not filled by direct access or through weighted lottery), higher education institutions were now allowed to use decentralized selective admissions (Goudappel, 1999; Kaiser & Vossensteyn, 2005, 2009). This was the moment, when the element of selective admissions, even though optional, was introduced into Dutch university admissions.

After these changes in admissions policies, it was possible to compare the study success of students admitted via lottery with the study success of students admitted according to selective admissions. The findings from a series of studies within one Dutch medical school indicated that the selected medical students performed better and dropped out less frequently than students who gained admissions based on the lottery (Urlings-Strop et al., 2009, 2011, 2013). Informed by the results of these studies, the Dutch government passed a law which abolished the weighted lottery system and introduced a selective system for oversubscribed programs. From 2017 onwards, it is a legal requirement to have at least two qualitative selection criteria for programs that do not follow an open admissions model but instead implement selective admissions (Higher Education and Scientific Research Act of the Netherlands [Wet Op Het Hoger Onderwijs en Wetenschappelijk Onderzoek], 2017).

1.4.2 The Bologna Process and Its Consequences for Graduate Selective Admissions

In 1987, a major bottom-up initiative emerged in the context of the European Union's policy initiatives in higher education, notably the successful ERASMUS program

(European Commission, 2017). It began as an exchange program for higher education students through the cooperative efforts of national authorities, HEIs, students, and has steadily evolved. The ERASMUS program has set the stage for more organized cooperation between HEIs (European Higher Education Area, n.d.).

In 1998, the higher education ministers of four European countries (France, Germany, the UK, and Italy) came together in Paris, France to discuss the external recognition of qualifications in higher education to further improve student mobility and employability. At this event, organized on the 800th anniversary of the Sorbonne University, the intention of creating an "open European area of higher education" was declared with the signing of the Sorbonne Joint Declaration (1998). The ministers from these four countries went on to encourage other Member States of the European Union to join them in the declared objectives. As a result, in 1999, Ministers of Education of 29 countries came together in Bologna, Italy and expressed their intention to commit to the development of the European Higher Education Area (EHEA), by signing the Bologna Declaration (1999). Since the signing, 48 countries and several European organizations are now participating in this intergovernmental higher education reform process, known as the Bologna Process. Every two to three years, Ministerial Conferences are held to discuss the advances within EHEA, make the agreements that facilitate comparability and quality of higher-education qualifications, and outline the EHEA's future actions.

One of the most significant changes, as announced in the Sorbonne Declaration and which the Bologna Process moved forward, concerned higher education degree structures in the participating countries. Before the start of the Bologna Process, there was a variety of national higher education degrees across Europe. For example, Dutch research universities, similar to their German counterparts, offered their students an integrated long first degree leading to the equivalent of a master's qualification. With the Bologna Process, a two-cycle degree structure (undergraduate and graduate, the latter leading to a master's and/or doctorate degree) was suggested for all participating countries. The Bologna degree structure was integrated in the Dutch Higher Education Act in 2002. In 2003, the two-cycle structure was finetuned into a three-cycle framework of qualifications (bachelor's/master's/PhD). Within this structure, bachelors' and masters' programs are considered as stages with an important option of transition between them (Teichler, 2007).

For the admissions field, the implementation of the Bologna degree structure resulted in two major changes. The first major change was that students with a bachelor's degree from universities of applied sciences, could enter masters' programs at Dutch research universities, provided these students fulfilled the requirements for the specific field. Before 2002, a certificate from a university of applied sciences did not offer a regular educational path to a research university (Witte et al., 2008). With the introduction of the Bologna degree structure, the number of students moving from universities of applied sciences to research universities increased substantially (see Figure 2). The influx of these students also meant the increased participation in research

universities of students from lower socioeconomic status (SES) backgrounds and certain minority groups (van de Werfhorst & van Tubergen, 2007).

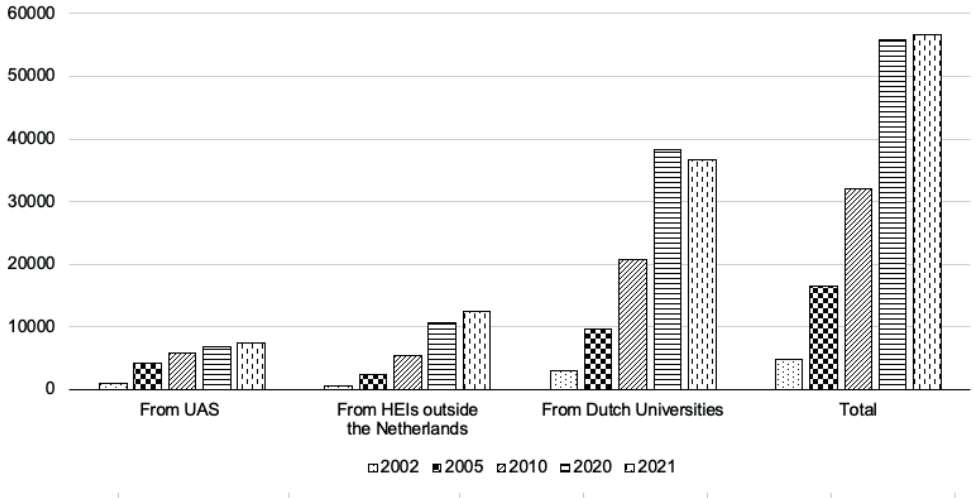


Figure 2 | *Inflow of First-Year Masters' Students at Dutch Research Universities over the Years 2002, 2005, 2010, 2020, and 2021.*

Note. *UAS stands for Dutch Universities of Applied Sciences; HEI stands for Higher Education Institution (Association of Universities in the Netherlands [VSNU], 2021).

The second major (and delayed) change was the implementation of an obligatory requirement to finish an undergraduate degree before entering a graduate program (so called "hard cut" or "harde knip"). It was introduced in the Dutch Higher Education Act only in 2012. On the one hand, "harde knip" aimed to contribute to limiting the time to complete a bachelor's degree. On the other hand, it aimed to ensure that bachelor's and master's programs are perceived as not combined but rather independent programs, which would facilitate students to make a (more) conscientious choice of a master's program (Stone, 2013) and increase mobility across professional and disciplinary fields as well as across borders. These two domestically quite delayed consequences of the implementation of the Bologna Process meant that the question—on which grounds to select students to masters' programs with limited places—became more relevant.

1.4.3 Increasing Internationalization of Dutch Higher Education

Another important Bologna-related development is the fact that Dutch universities have been receiving increasing numbers of applications from international students (see Figure 2). In 2021–2022, international students constitute up to 23% of the total university student population (Universiteiten van Nederland, 2022b). For university freshmen, these numbers are even higher: Around 40% of university freshman are international (Centraal Bureau voor de Statistiek; 2022). Most of international students (73%) come from within the European Economic Area (Universiteiten van Nederland, 2022b).

The Dutch universities have become increasingly more attractive to international students in recent years due to the abundance of English-taught programs at master's level (with 74% of masters' programs being English-taught; Association of Universities in the Netherlands, 2018) and the success of all 13 Dutch research universities in the university rankings (e.g., all 13 Dutch research universities are in top 250 according to the Times Higher Education (2022).

This increase from students with diverse education backgrounds requires enhanced professional knowledge and quality assurance of universities' selective admissions systems. It is obviously more complex to assess growing diversity in application files compared to a situation when students transfer within the same higher education institution or come from the same national higher educational system.

1.5 (Graduate) Selective Admissions and Equity Considerations

Along with increased number and diversified body of masters' students at Dutch research universities over the last 20 years, expectations regarding fairness, inclusiveness, and transparency have also been on the rise due to concerns around growing socioeconomic inequality (Haveman & Smeeding, 2006; Piketty, 2013). Interestingly, equity (and its dimensions such as "equity for equal needs", "equity for equal potential", and "equity for equal achievement"; Espinoza, 2007) and equality (and its dimensions "equality of opportunity", "equality for all", and "equality on average across social groups"; Espinoza, 2007) were not prioritized in the higher educational policies of most Western European countries until the 1990s (Teichler, 2007). In this regard, the major policy development was the Sorbonne Joint Declaration (1998), that stated that "students should be able to enter the academic world at any time in their professional life and from diverse backgrounds" (Sorbonne Joint Declaration, 1998, p. 2). The countries that joined the Bologna Process (including the Netherlands) reaffirmed their commitment to the Sorbonne Declaration ensuring diversity within the student bodies of their higher education institutions.

Most of the research on widening participation in higher education of underrepresented groups has been conducted unsurprisingly on the undergraduate level (see for example Torotcoi et al., 2020 for a review of efficiency of measures to enhance access to higher education across European countries). This is because of the crucial importance the transition from secondary school to higher education plays in providing students from underrepresented backgrounds a chance of entering the trajectory of higher education. However, rarely is social stratification and the reinforcing or reducing of inequality in graduate education recognized (Posselt & Grodsky, 2017). Yet, demand for access to graduate education is increasing (Payne, 2015). Therefore, the graduate admissions systems should also safeguard the quality of process, decisions, and outcomes and account for equity considerations. Quality assurance implies, among

others, that the admissions process is not prone to admissions biases, ensures a diverse student population, and does not induce students being rejected who would have otherwise been successful if they had been admitted, known as "false negatives" (a famous "false negative" case is presented in Box 1).

Box 1

A Famous Mistake of University Admissions: Wilhelm Conrad Röntgen

A historical example that represents a classic "false negative" case in admissions is the case of Wilhelm Conrad Röntgen (1845–1923): a mechanical engineer and physicist and the first Nobel Prize laureate in Physics for his discovery of X-rays (in many languages, X-radiation is referred to as "Röntgen radiation"). At the age of 20, Röntgen was not admitted to Rijksuniversiteit Utrecht (now Utrecht University), the Netherlands, at the start of his academic studies due to specific local admissions requirements and thus attended the university only as a visitor. For enrolling as a regular student, the university required a higher secondary school diploma, which he did not obtain (according to rumours, one of his school teachers took a dislike to him). Röntgen learned, however, that the Federal Polytechnic Institute in Zurich (today known as ETH Zurich) admitted students based on university entrance examinations, even without formally possessing a higher secondary school diploma. He became a student there and graduated with a PhD from the University of Zurich. During his outstanding research career, he took appointments at several European universities, including University of Strasbourg, University of Hohenheim, University of Giessen, University of Würzburg, and University of Munich. Being offered a professorship position at Rijksuniversiteit Utrecht at the peak of this research career, he refused it (Utrecht University, n.d.).

Research on "false negative" cases is extremely scarce because the information on the study success of these students is rarely available; being rejected, their study success on other educational programs is hard to follow, while it is virtually impossible to assess their hypothetical performance on a program to which they were not admitted. The inferences on selection decisions, accounting for "false negative" cases are possible to make only in exceptional situations, when (almost) all applicants are admitted (Kurysheva et al., 2022; van der Linden, 2018; van Ooijen-van der Linden et al., 2017). In the data collected for this thesis, the information on "false negative" cases was not available.

In summary, guided by (1) the abolishment of the weighted lottery approach and the introduction of selective admissions, (2) increasing numbers of (inter)national applicants, (3) amplified diversification of application files and a commitment to diversity in the student body, and (4) societal expectations for fair, objective, inclusive,

and transparent admissions, the administrations of Dutch research universities search for evidence-based insights on implementation of selective admissions processes in general and for admissions to masters' programs in particular.

1.6 The Scope and Rationale of This Thesis

1.6.1 Theoretical Perspectives

This thesis is written from the perspective of psychology of individual differences in aptitude and learning, also known as differential psychology (Stern, 1900; Tyler, 1965). Employing quantitative methods, psychology of individual differences aims to study (a) nature and extent of individual differences and (b) interrelationships of mental processes (Tyler, 1965). The perspective of differential psychology has been employed in this thesis because it most closely relates to the main purpose of this work: to establish the interrelations of individual-level variables (in particular, the interrelations between undergraduate academic indicators and graduate study success). Moreover, admissions committees make their decisions, using individual-level data about student characteristics, including various measurements of student abilities (such as grades, standardized tests, recommendation letters etc.). From this perspective, differential psychology is again a relevant disciplinary field because it primarily focuses on measurements of individual abilities.

Research on the selection of individuals has been conducted utilizing various other theoretical perspectives, including evolutionary biology (Darwin, 1859), sociology (Mountford Zimdars, 2016; Posselt, 2016; Posselt & Grodsky, 2017; Wakeling, 2021; Warikoo, 2016), and human capital theory (López-Cabrales et al., 2011; Polachek, 1981). They have also been analyzed from a historical (Karabel, 2006; Douglass, 2010, amongst others) and political philosophy point of view (e.g., Sandel, 2020). These theoretical perspectives consider other aspects of selection of individuals such as differences in phenotype[2], the context of inequality in which selection takes place (foremost, socio-economic diversity of applicants, the challenges of meritocratic idea for equal access to higher education, social stratification, cognitive biases emerging during individual and group decision making, educational and occupational self-selection, etc.). While recognizing their importance and referring to some of these aspects at various points in this thesis, they do not constitute the primary scope of the present work, which takes a psychological perspective.

2 Phenotype is "the observable characteristics or traits of an organism that are produced by the interaction of the genotype and the environment: the physical expression of one or more genes" (Merriam-Webster dictionary, n.d.)

1.6.2 A Focus on Life and Natural Sciences (as a Part of the STEM Disciplines)

This thesis focuses on selection methods for the life and natural sciences graduate programs. Both these fields can be subsumed under the broader term Science, Technology, Engineering, and Mathematics (STEM). Overall, STEM fields play an important role in technological innovation and job creation (Council of Graduate Schools, 2013) and thus benefit the development of society.

Selective admissions are particularly relevant for the life and natural sciences graduate programs. These programs are often characterized by student participation in research internships, which usually happens in academic research groups and involves the usage of different kind of (costly) research materials. This limits the number of study places on these programs, in addition to the limitations imposed by other factors that also hold for a more course-oriented curriculum (e.g., a limited number of teachers) and the factors described in sections 1.4—1.5 (i.e., diversification of application files, equity considerations, increasing number of students: the inflow of first-year masters' students in life and natural sciences at research universities in the Netherlands doubled over the last 10 years: from 4,004 in 2012 up to 8,134 in 2021; Universiteiten van Nederland, 2022a).

Aiming to validate the selection methods used at life and natural sciences programs, we use the data of students on these programs in our empirical studies (Chapter 3, 4, and 5). In our review (Chapter 2), we included the primary studies whose data sets had comprised students in all STEM fields (which means also, for example, engineering students). It was not feasible to disentangle the findings on students in the life and natural sciences from findings on other STEM students, however, we do not regard this as a substantial issue, because the selection methods used in different STEM fields are rather comparable.

1.6.3 Problem Statement. Theoretical and Practical Relevance of This Research

To date, there has been no comprehensive review of the methods for selection in admissions to graduate studies. Reviews have mostly been conducted within the medical education field or the fields of undergraduate and personnel selection. Therefore, a mapping review of the existing knowledge on methods for graduate level selection can enhance our understating of the current scientific consensus on which selection methods are evidence-based specifically for admissions to a graduate level.

Most studies on selection methods in graduate admissions are conducted in the USA. There is a lack of research on this topic in the European context, especially for the Dutch research masters', which were the first type of graduate programs after the introduction of the Bologna degree structure that were allowed to select students. Considering the factors surrounding the admissions debate which are specific to the Dutch higher education context, the findings from other education systems cannot

be easily generalized[3] to the Dutch context. Conducting empirical studies within the context of the Dutch research masters' allows us to fill this gap and to provide more insights on Dutch (and thus European) graduate admissions systems.

Furthermore, some selection methods and graduate study success dimensions have not yet been sufficiently addressed in the field of graduate admissions. For example, the examination of research-related predictors (e.g., grade for bachelor's thesis) and research-related outcomes (e.g., grade for a research internship, assessments on rubrics of research skills, and research report) have not been explored in prior studies. Our focus on research masters' allows us to tackle these aspects and shed some light on their usefulness in admissions. Despite the focus on STEM disciplines (and predominantly, life/natural sciences), it is hoped that insights can be found that are also relevant for graduate schools at large (i.e., in other disciplines), but this depends on the rate of generalization possible.

From the practical side, the increasing competitiveness for graduate study places in conjunction with the decentralized nature of graduate selective admissions necessitates renewed scientific attention to selective admissions procedures. This challenge is becoming more prominent for many Dutch graduate schools because they often use decentralized admissions, where predominantly academic staff committees are responsible for selection. The practical intention of this thesis is to provide insights on valid selection methods for assisting academic staff committees with their admissions work.

1.6.4 Definition of Study Success

In the empirical studies reported in this thesis, it was chosen to operationalize graduate study success through outcomes within the timeline of a graduate educational program. Namely, study success is operationalized through five dimensions: (1) graduate degree attainment, (2) Graduate Grade Point Average (GGPA), (3) grade for research internships (the essential part of a graduate research training), (4) supervisors' scores on performing research internships, and (5) time to graduate degree[4]. This implies that the efficacy

3 One of the reasons for this generalization difficulty is because the extent of Bologna-related harmonization appears to be less notable than had been foreseen (Teichler, 2007). In other words, not all structural differences between higher education systems dissolved with the introduction of the Bologna Process as there are many exceptions and specific cases. For example, the exceptions exist from the fundamental model of 3+2 (i.e., three years of a bachelor's program and two years of a master's program; Teichler, 2007; Witte, 2008). Moreover, in addition to European-level reforms and global developments, structural differences in higher education systems stay responsive to and are partly shaped by local political and economic events (Palfreyman & Tapper, 2009). On the other hand, Bologna-related harmonization of systems provokes enhanced mobility (Vögtle, 2019). Therefore, studying selective admissions within the Dutch context is relevant for understanding the developments in this area within the European context.

4 The choice was made to use time to degree, similar to some studies in the field and not to use study progress as some other studies do (e.g., Niessen, 2018; Zimmermann et al., 2018). Study progress (sometimes also referred as rate of progress) is the number of credits obtained in a master's program divided by the number of study semesters until completion. The choice was made to use time to graduate degree as one of the study success measures because study progress is not a sensible outcome measure at the specific

of selection methods toward outcomes which are observed beyond graduate studies (such as job attainment, job performance, or being active citizens) are outside of the scope of this thesis. However, a mapping review, which is part of this thesis, presents research evidence toward all graduate outcomes that have been examined in the relevant studies across the last 15 years, though the majority of them do not go beyond the timeline of graduate programs.

1.7 The Design of This Thesis

1.7.1 Aims

Based on the theoretical and practical considerations outlined above, the overarching purpose of this research project is to enrich knowledge on methods for evidenced-based, fair, and transparent graduate student selective admissions to masters' programs in STEM (and predominantly, life/natural sciences)[5] disciplines within the Dutch higher education context. The first aim of this thesis is to explore the evidence on different graduate selection methods in STEM fields. The second aim is to explore and evaluate the current selection criteria and methods for STEM masters' programs at a major Dutch research university. The third aim is to determine the potential predictive values of the selection methods for graduate study success which are not integrated into the current selection practice for most Dutch STEM research masters' programs.

These aims can also be viewed as three perspectives: retrospective (looking backwards to assess the extent to which the applied selection methods have been shown in prior research to predict study success), contemporary (to establish the current state of selective admissions practices), and prospective (to pilot methods that would better meet the criteria of objective, transparent, and fair selection methods).

It is hoped that the "selection toolkit" potentially resulting from this research would be used to assist graduate schools' administration and admissions committees in designing a sound admissions process. A process should strive to form a diverse and international population of potentially successful students, while also complying with relevant legislation and contributing to the aims of the higher education system.

1.7.2 Research Questions

1) What evidence is provided in the research literature on the extent to what different selection methods in graduate admissions in STEM fields are valid, reliable,

masters' programs, which were examined within this thesis. At the Graduate School of Life Sciences, where the empirical studies were conducted, students often get their first grade for an internship of 51 EC in 9 to 12 months after the start of their graduate program. Therefore, study progress per month or even per semester would not be a sensible operationalization of study success.

5 Having clarified the scope in section 1.6.2., we will further refer to these programs as simply STEM programs

accepted by the stakeholders, and cost-effective? What are the procedural issues of the existing selection methods in graduate admissions?

2) What kind of selection criteria and methods in graduate admissions are reported to be important in decision making for admissions to STEM masters' programs by responsible staff? How frequently are they applied? Are they transparent to the applicants of these masters' programs?
3) To what extent do different undergraduate academic indicators, available in institutional admissions data, predict different dimensions of graduate study success (e.g., graduate degree completion, GGPA, graduate major research project grade, and graduate time to degree)?
4) To what extent could the Graduate Record Examinations (GRE) General test predict the study outcomes on research-oriented graduate programs (such as level of research skills and grade for a major research project), while taking into the account the effects of socioeconomic status?

1.7.3 Thesis Outline

Figure 3 provides a schematic overview of the steps, which are undertaken in this thesis in order to answer the research questions presented above.

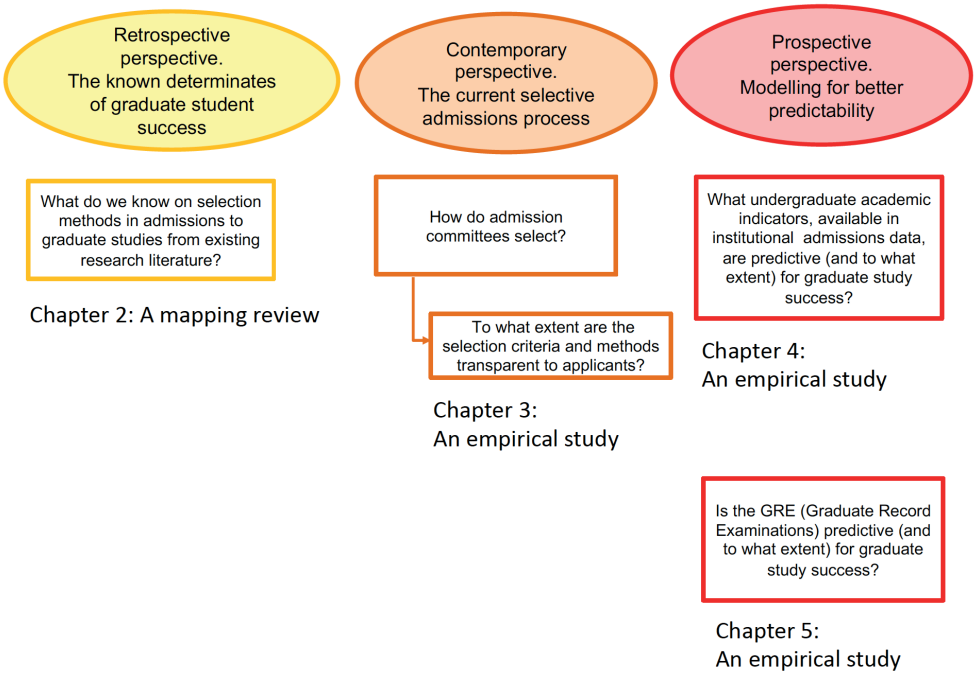


Figure 3 | *An Overview of the Perspectives Undertaken in the Present Thesis.*

For exploring evidence on different graduate selection methods (aim 1), a mapping review was conducted (Chapter 2). Using a systematic search of research literature over the last 15 years, the review maps the variety of selection methods for admissions of STEM graduate students. It compares each selection method against four evaluative quality principles: (1) predictive validity and reliability, (2) procedural issues, (3) acceptability, and (4) cost-effectiveness.

For exploring and evaluating the current admissions process (aim 2), an empirical survey study was conducted (Chapter 3). This chapter distinguishes between selection criteria and selection methods and explores their application in practice by admissions committees of various STEM research masters' programs. In addition, it examines to what extent the applied selection criteria and methods are transparent to the applicants. The data for this survey study were gathered at two large graduate schools within one Dutch research university.

For detecting graduate selection methods that are potentially valid but are not integrated into the current selection practice of STEM research masters' programs (aim 3), an empirical study was conducted (Chapter 4). It explores predictive validity of several undergraduate academic indicators for graduate study success using institutional data of students in the life sciences. With the same aim, another empirical study was conducted, which assesses the predictive validity of the Graduate Record Examinations (GRE) on a sample of research masters' students from three STEM graduate schools at one Dutch research university (Chapter 5).

1.8 References

Accreditation Organization of the Netherlands and Flanders. (2016). *Specification of additional criteria for research master's programmes*. https://www.nvao.net/files/attachments/.1255/Nadere_uitwerking_aanvullende_criteria_onderzoeksmasters_30_mei_2016_EN_def_nieuwe_huisstijl.pdf

Association of Universities in the Netherlands. (2021). *Instroom nieuwe masterstudenten [Influx new masters' students]*. https://vsnu.nl/instroom-master.html

Association of Universities in the Netherlands. (2018, May 14th). *Internationalization agenda for higher education*. Netherlands Association of Universities of Applied Sciences. Association of Universities in the Netherlands. https://www.universiteitenvannederland.nl/files/documents/Internationalisation%20Agenda%20for%20Higher%20Education.pdf

Centraal Bureau voor de Statistiek (2022). 40 procent eerstejaars universiteit is internationale student [40 percent freshman university is international student]. https://www.cbs.nl/nl-nl/nieuws/2022/11/40-procent-eerstejaars-universiteit-is-internationale-student

Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. John Murray.

Douglass J. A. (2010). From chaos to order and back? A revisionist reflection on the California Master Plan for Higher Education@50 and thoughts about its future. Research & Occasional Paper Series: CSHE.7.10. https://escholarship.org/uc/item/6q49t0hj

Education GPS, Organization for Economic Co-operation and Development (OECD). (2022). Copyright 2020 by Education GPS. https://gpseducation.oecd.org/CountryProfile?primaryCountry=NLD

European Commission. (2017). *From Erasmus to Erasmus+: a story of 30 years.* https://ec.europa.eu/commission/presscorner/detail/en/MEMO_17_83

European Higher Education Area (n.d.). *EU Support on Bologna Commitments*. http://ehea.info/page-european-commission

Espinoza, O. (2007). Solving the equity–equality conceptual dilemma: A new model for analysis of the educational process. *Educational Research*, *49*(4), 343–363. https://doi.org/10.1080/00131880701717198

Goudappel, F. (1999). The Dutch system of lottery for studies. *European Journal for Education Law and Policy*, *3*, 23–27. https://eur.academia.edu/FloraGoudappel

Haveman, R., & Smeeding, T. (2006). The role of higher education in social mobility. *The Future of Children*, *16*(2), 125–150. https://www.jstor.org/stable/3844794

Higher education and scientific research Act of the Netherlands [Wet op het hoger onderwijs en wetenschappelijk onderzoek], § 7.30b (2017). https://wetten.overheid.nl/BWBR0005682/2017-09-01

Kaiser, F., & Vossensteyn, H. (2005). Access to Dutch higher education. In T. Tapper & D. Palfreyman (Eds.), *Understanding mass higher education. Comparative perspeictives on access* (pp. 92–120). Routledge.

Kaiser, F., & Vossensteyn, H. (2009). Excellence in Dutch higher education: Handle with care. In D. Palfreyman & T. Tapper (Eds.), *Structuring mass higher education. The role of elite institutions* (pp. 169–181). Routledge.

Karabel, J. (2006). *The chosen: The hidden history of admission and exclusion at Harvard, Yale, and Princeton*. Houghton Mifflin.

Kurysheva, A., van Ooijen-van der Linden, L., van der Smagt, M., & Dilaver, G. (2022). Potential and unique incremental value of signal detection theory as a research method in higher education: An example from selective graduate admissions [Manuscript submitted for publication]. Center of Education and Training, University Medical Center Utrecht & Faculty of Social and Behavioural Sciences, Utrecht University.

López-Cabrales, Á., Real, J. C., & Valle, R. (2011). Relationships between human resource management practices and organizational learning capability: The mediating role of human capital. *Personnel Review*, *40*(3), 344–363. https://doi.org/10.1108/00483481111118658

Merriam-Webster dictionary. (n.d.). Phenotype. In *Merriam-Webster.com dictionary*. Retrieved January 31, 2022, from https://www.merriam-webster.com/dictionary/phenotype

Mountford Zimdars A. (2016). *Meritocracy and the university: Selective admission in England and the United States*. Bloomsbury Academic.

Niessen, S. (2018). *New rules, new tools: Predicting academic achievement in college admissions* [Doctoral dissertation, University of Groningen]. Rijksuniversiteit Groningen. https://research.rug.nl/en/publications/new-rules-new-tools-predicting-academic-achievement-in-college-ad

Palfreyman, D., & Tapper, T. (2009). Preface. In D. Palfreyman & T. Tapper (Eds.), *Structuring mass higher education. The role of elite institutions* (pp. xxi–xxii). Routledge.

Payne, D. (2015). A common yardstick for graduate education in Europe. *Journal of the European Higher Education Area*, *2*, 21–48. https://www.handbook-internationalisation.com/en/handbuch/gliederung/?articleID=566#/Beitragsdetailansicht/675/566

Pinxten M., de Laet T., van Soom C., & Langie G. (2015). *Fighting increasing dropout rates in the STEM field: The European readySTEMgo project* [Paper presentation]. 43rd Annual SEFI Conference 2015, Orléans, France. https://www.sefi.be/wp-content/uploads/2017/09/54365.-M.PINXTEN.pdf

Piketty, T. (2013). *Capital in the twenty-first century*. The Belknap Press of Harvard University Press.

Polachek, S. W. (1981). Occupational self-selection: A human capital approach to sex differences in occupational structure. *The Review of Economics and Statistics*, *63*(1), 60–69. https://doi.org/10.2307/1924218

Posselt, J. R. (2016). *Inside graduate admissions. Merit, diversity, and faculty gatekeeping*. Harvard University Press.

Posselt, J. R., & Grodsky, E. (2017). Graduate education and social stratification. *Annual Review of Sociology*, *43*, 353–378. https://doi.org/10.1146/annurev-soc-081715-074324

Sandel, M. J. (2020). *The tyranny of merit. What's become of the common good?* Farrar, Straus and Giroux.

Schripsema, N. R. (2017). *Medical student selection: Effects of different admissions processes* [Doctoral dissertation, University of Groningen]. Rijksuniversiteit Groningen. https://research.rug.nl/en/publications/medical-student-selection-effects-of-different-admissions-process

Snijder, J. K. (2016). *Mastering research: A study about the effectiveness of research master programmes in the Netherlands* [Doctoral dissertation, University of Twente]. Universiteit Twente. https://research.utwente.nl/en/publications/mastering-research-a-study-about-the-effectiveness-of-research-ma

Sorbonne Joint Declaration. Joint declaration on harmonisation of the architecture of the European higher education system, May 25, 1998. http://ehea.info/page-sorbonne-declaration-1998

Stern, W. (1900). *Über Psychologie der individuellen Differenzen: Ideen zu einer „differentiellen Psychologie"* [On the psychology of individual differences: Toward a „differential psychology"]. Barth.

Stone, P. (2013). Access to higher education by the luck of the draw. *Comparative Education Review*, *57*(3), 577–599. https://doi.org/10.1086/670663

Tapper, T., & Palfreyman, D. (2009). Structuring mass higher education: Interpreting the process of change. In D. Palfreyman & T. Tapper (Eds.), *Structuring mass higher Education. The role of elite institutions* (pp. 1–9). Routledge.

Teichler, U. (2007). *Higher education systems: Conceptual frameworks, comparative perspectives, empirical findings* (Vol. 9). Sense Publishers.

The Bologna Declaration. Joint declaration of the European Ministers of Education, June 19, 1999. http://www.ehea.info/page-ministerial-conference-bologna-1999

The Times Higher Education. (2022). *World university rankings 2022*. https://www.timeshighereducation.com/world-university-rankings/2021/world-ranking#!/page/0/length/25/sort_by/rank/sort_order/asc/cols/stats

Torotcoi, S., Gologan, D., & Kurysheva, A. (2020). What works for underrepresented groups? Identifying effective measures for enhancing access to higher education in Europe. In A. Curaj, L. Deca, & R. Pricopie (Eds.), *European Higher Education Area: Challenges for a new decade* (pp. 177–196). Springer International Publishing. https://doi.org/10.1007/978-3-030-56316-5_13

Tyler, L. (1965). *The psychology of human differences*. Appleton-Century-Crofts, Division of Meredith Publishing Company.

Universiteiten van Nederland. (2022a). Intake of new Master's students. https://www.universiteitenvannederland.nl/en_GB/instroom-master.html

Universiteiten van Nederland. (2022b). Internationale studenten in het wetenschappelijk onderwijs [International students in Science Education]. https://www.universiteitenvannederland.nl/f_c_internationale_studenten.html

Urlings-Strop, L. C., Stegers-Jager, K. M., Stijnen, T., & Themmen, A. P. N. (2013). Academic and non-academic selection criteria in predicting medical school performance. *Medical Teacher*, *35*(6), 497–502. https://doi.org/10.3109/0142159X.2013.774333

Urlings-Strop, L. C., Stijnen, T., Themmen, A. P. N., & Splinter, T. A. W. (2009). Selection of medical students: A controlled experiment. *Medical Education*, *43*(2), 175–183. https://doi.org/10.1111/j.1365-2923.2008.03267.x

Urlings-Strop, L. C., Themmen, A. P. N., Stijnen, T., & Splinter, T. A. W. (2011). Selected medical students achieve better than lottery-admitted students during clerkships: Achievement of selected and lottery-admitted medical students. *Medical Education*, *45*(10), 1032–1040. https://doi.org/10.1111/j.1365-2923.2011.04031.x

Utrecht University. (n.d.). *Wilhelm Conrad Röntgen*. https://www.uu.nl/en/organisation/wilhelm-conrad-rontgen

van de Werfhorst, H. G., & van Tubergen, F. (2007). Ethnicity, schooling, and merit in the Netherlands. *Ethnicities*, *7*(3), 416–444. https://doi.org/10.1177/1468796807080236

van der Linden, L. (2018). *Prediction of study success - creation of magic zones* [Doctoral dissertation, Utrecht University]. https://www.narcis.nl/publication/RecordID/oai:dspace.library.uu.nl:1874%2F373433

van der Wende, M. C. (2020). Higher education in the Netherlands. In *The SAGE encyclopedia of higher education*. SAGE Publications.

van Ooijen-van der Linden, L., van der Smagt, M. J., Woertman, L., & te Pas, S. F. (2017). Signal detection theory as a tool for successful student selection. *Assessment & Evaluation in Higher Education*, *42*(8), 1193–1207. https://doi.org/10.1080/02602938.2016.1241860

Vögtle, E. M. (2019). 20 years of Bologna - A story of success, a story of failure. *Innovation: The European Journal of Social Science Research*, *32*(4), 406–428. https://doi.org/10.1080/13511610.2019.1594717

Wakeling, P., & Mateos-González, J. L. (2021). *Inequality in the highest degree*. *Postgraduates, prices and participation*. the Sutton Trust. https://dera.ioe.ac.uk/38128/1/Inequality-in-the-Highest-Degree-Final-Report.pdf

Warikoo, N. K. (2016). *The diversity bargain and other dilemmas of race, admissions, and meritocracy at elite universities*. The University of Chicago Press.

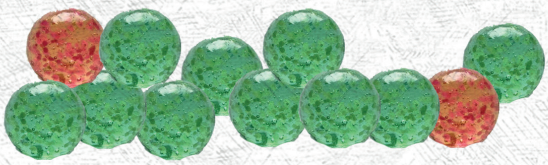
Williams, K. M., & Wendler, C. (2020). The open admissions model: An example from the United States. In *Higher education admission practices: An international perspective*. Cambridge University Press.

Witte, J., van der Wende, M., & Huisman, J. (2008). Blurring boundaries: How the Bologna process changes the relationship between university and non-university higher education in Germany, the Netherlands and France. *Studies in Higher Education*, *33*(3), 217–231. https://doi.org/10.1080/03075070802049129

Zimmermann, J., von Davier, A. A., Buhmann, J. M., & Heinimann, H. R. (2017). Validity of GRE General test scores and TOEFL scores for graduate admission to a technical university in Western Europe. *European Journal of Engineering Education*, *43*(1), 144–165. https://doi.org/10.1080/03043797.2017.1343277

Zwick, R. (2019). Fairness in measurement and selection: Statistical, philosophical, and public perspectives. *Educational Measurement: Issues and Practice*, *38*(4), 34–41. https://doi.org/10.1111/emip.12299

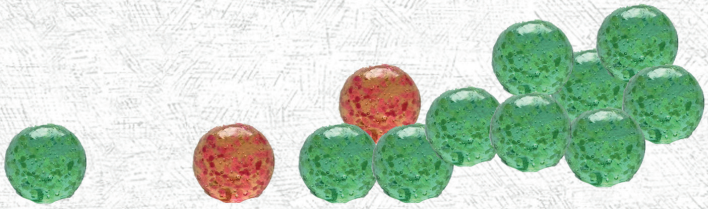
2



Acknowledgements

The author is grateful to Cecily Stolte for acting as second coder during both steps of literature screening. The author is also thankful to Adam Frick and Christine Fox for copy editing.

Validity, Acceptability, and Procedural Issues of Selection Methods for Graduate Study Admissions in the Fields of Science, Technology, Engineering, and Mathematics: A Mapping Review



Author note

An adapted version of this chapter is in preparation for submission to a peer-reviewed journal as "Kurysheva, A. (2022). *Validity, Acceptability, and Procedural Issues of Selection Methods for Graduate Study Admissions in the Fields of Science, Technology, Engineering, and Mathematics: A Mapping Review* [Manuscript in preparation]".

Abstract

This review presents the first comprehensive synthesis of available research on both cognitive and noncognitive selection methods for graduate study admissions. It focuses on STEM disciplines and covers the period between 2005 and 2020. Ten categories of graduate selection methods emerged. Each category was critically appraised against the following evaluative quality principles: predictive validity and reliability, acceptability, procedural issues, and cost-effectiveness. Gaps and further directions in research literature were identified. Theoretical, practical, and social implications of using (non-)evidence-based selection methods are discussed. Overall, this synthesis of the latest findings in the field of graduate selective admissions allows admissions committees to choose which selection methods to use and which essential aspects of their implementation to account for.

Keywords: review, graduate admissions, selection methods, predictive validity, acceptability.

2.1 Introduction

A high-quality student selection procedure for graduate level education is of utmost importance for programs, students, and society. Higher education has seen several influential policy developments over the past decades such as the introduction of the Bologna Process in 1999 in Europe or the increased internationalization of higher education across the globe. These policies contributed to rising international/cross-border and national (i.e., between higher education institutions within one country) student mobility (Okahana & Zhou, 2018; Payne, 2015). The knock-on effect of this mobility has created a growing diversity of graduate application files. Admissions committees are now faced with applicants from different higher education systems, potentially a variety of background fields, and varying levels of academic skills and proficiency in the language of instruction.

Furthermore, the problem of underrepresentation of students with certain backgrounds persists across the globe, including countries with well-developed higher education systems (Salmi & Bassett, 2014). As such, it is still harder for students with low socioeconomic status (SES), a migrant background or of a certain race, first-generation students, or students with disabilities to get into higher education programs (Garaz & Torotcoi, 2017; Salmi & Bassett, 2014; Weedon, 2017). Students' application files are often conditioned by their background: For example, students with parents of low SES cannot typically show an impressive list of extracurricular activities on their resume in contrast to their peers with parents of high SES (Jayakumar & Page, 2021). It is, therefore, often the case that a straightforward assessment of application files is not feasible because of the multifaceted nature of each application. Unsurprisingly, it is a complex task for admissions committees to evaluate the educational background and achievements of (inter)national students with diverse backgrounds. Regardless of described complexities, admissions decisions must be objective, fair, and transparent to ensure their adequate justification.

To facilitate the achievement of these overarching goals, four evaluative quality principles[6] regarding student selection methods were outlined as essential (Patterson et al., 2016). These are: (1) validity and reliability, (2) procedural issues, (3) acceptability, and (4) cost-effectiveness. These evaluative quality principles generally aim to answer the following respective questions: (1) whether a selection method works, and whether it works better than other selection methods; (2) what are the concerns and limitations of the selection method; (3) how widely is the selection method applied, to what extent are admissions committees willing to apply the method, and how much is the selection method favored by applicants; (4) what is the financial effect of the selection method on educational programs and applicants (Patterson et al., 2016).

6 In the original article of Patterson et al. (2016), they are called "evaluative quality criteria". We call them "evaluative quality principles" as the usage of the term "criteria" might confuse readers with "selection criteria" in Chapter 3 of this thesis.

There is a striking lack of studies that synthesize research evidence on selection methods for graduate study admissions while accounting for all four evaluative quality principles. Instead, the existing reviews and meta-analyses address evidence for each selection method separately: standardized testing (Kuncel et al., 2004, 2010; Kuncel & Hezlett, 2007b, 2010), recommendation letters (Kuncel et al., 2014), personal statements (Murphy et al., 2009), and other various noncognitive measures (Kuncel et al., 2020; Kyllonen et al, 2005, 2011; Megginson, 2009). Moreover, these studies usually focus on predictive validity and rarely on procedural issues, with only limited or no attention to reliability, acceptability, and cost-effectiveness.

The only review to combine evidence on all available selection methods within one study and included the four evaluative quality principles (validity/reliability, procedural issues, acceptability, and cost-effectiveness) was conducted by Patterson et al. (2016). However, this review only focused on selection methods in medical education. For example, it does not present evidence on (nonmedical) standardized tests of academic aptitude, tests of language of instruction, or amount and quality of prior research experience. Therefore, its findings can only be partially generalized for graduate admissions.

The question that arises is which educational field (except medical education) has attracted enough high-quality research that (a) addresses the four evaluative quality principles and (b) allows admissions committees to use the findings in a wide range of graduate programs, therefore, enhancing the potential impact of this review? From the preliminary overview, we think that science, technology, engineering, and mathematics (STEM) fields meet these two conditions. STEM fields have been recognized worldwide as fundamental for finding solutions to urgent societal problems (Proudfoot & Hoffer, 2016). The efforts of certain countries to become leaders in STEM higher education and research (e.g., China; Kirby & van der Wende, 2019) are illustrative of how crucial the STEM fields are for economic growth and prosperity. Unsurprisingly, STEM disciplines have attracted a rising number of students, making research evidence on selection methods for STEM studies increasingly more relevant. Since there has been no synthesis of such evidence to date, we designed this review to address this gap.

2.1.1 The Present Review

The aim of this review is to present a comprehensive overview of research evidence on the existing selection methods in graduate admissions in STEM fields. The review focuses on evaluative quality principles of validity, reliability, procedural issues, acceptability, and cost-effectiveness. The term "graduate" refers to both master's and doctoral levels. That is, studies on both levels were collected for this review.

2.1.1.1 Research Questions

What evidence is provided in research literature within STEM graduate admissions field on:

1) the extent to which different selection methods are valid and reliable?
2) procedural issues of the selection methods?
3) the extent to which different selection methods are accepted by stakeholders?
4) the extent to which different selection methods are cost-effective?

2.2 Method

For this review, a systematic search was conducted and complemented with an expanded search of literature in reference lists of relevant books and articles.

2.2.1 Inclusion Criteria for the Literature Review

The inclusion criteria for this review were:

- the topic on selection methods in graduate admissions,
- the graduate level of education (i.e., master's and/or PhD phase),
- samples that include students from STEM disciplines,
- studies addressing at least one of four evaluative quality principles of interest: validity/reliability, procedural issues, acceptability, and cost-effectiveness,
- studies conducted in at least one of the Organization for Economic Co-operation and Development (OECD) countries[7],
- studies published in English,
- studies published[8] in scientific journals, edited books, and doctoral dissertations,
- studies conducted in the period between 2005–2020.

The time frame was chosen in accordance with the changes in European higher education systems after the introduction of the Bologna Process (The Bologna Declaration, 1999). Countries joined the process in different subsequent years. Therefore, 2005 was chosen as a plausible cut-off moment to account for the fact that the first students, studying within the new system, could graduate. The same time frame was applied for the US research context.

We chose to review the literature, referring to master's and PhD levels together (that is, on a graduate level overall) because the training on both levels is advanced. Furthermore, many studies that were included in this review did not make a distinction between the two levels. We also considered different STEM majors or contexts (e.g., the European vs. the US contexts) together because we aimed to detect overarching

7 The OECD countries were chosen because of their well-developed higher education systems as well as an expectation that the quality of research in these countries is comparable

8 This choice was made to avoid unverified results and statements from gray literature.

patterns in evaluative quality principles that would be applicable to a variety of majors and higher education contexts on a graduate level.

2.2.2 The Literature Search Procedure

The literature search delivered 3,230 potentially relevant items including duplicates. The main portion of the results was obtained via conducting a systematic search in a specialized databases (ERIC: n = 1,089; PsycInfo: n = 1,112; Medline: n = 234; Scopus: n = 649). The keywords of the systematic search can be found in Table A1 in the Appendix. The syntax per database is available upon request. Next, the literature search was extended beyond the database approaches. Namely, the citations from relevant articles were examined (n = 62), and previously collected research literature was added (n = 84). The screening was conducted in two steps. In the first step, the titles and abstracts were scanned to remove duplicates and obvious irrelevant search results. In the second step, the full texts of remaining articles were obtained[9] and examined.

Figure 1 presents a detailed flowchart of the steps undertaken. Two coders (the author and a research assistant) conducted both steps of screenings. The agreement after the first screening was near complete agreement (kappa = .88) and that of the second screening was strong agreement (kappa = .70). All the disagreements were resolved in discussion. In total, 80 articles met the inclusion criteria for this review. The distribution across the OECD countries is presented in Table 1.

Table 1 | *Distribution of the Studies Included in the Review across the OECD Countries.*

OECD Country	Number of studies included in this review
Not specified/across continents	19
Across Europe	2
Belgium	1
Canada	1
Mexico	1
The Netherlands	4
Puerto Rico (the unincorporated territory of the US)	1
Switzerland	2
Turkey	1
The US	48
Total number of articles	80

After the screening was completed, the 80 studies were assigned into ten graduate selection method categories: (1) prior grades, (2) standardized testing of academic abilities, (3) letters of recommendation, (4) interviews, (5) personal statements (i.e., motivation letters), (6) personality assessments, (7) intelligence assessments, (8)

9 The full texts of four articles were not found even after contacting authors and were not included in the final number.

language proficiency, (9) prior research experience, and (10) various, rarely studied selection methods that do not fall under more common methods above (such as resumes, selectivity of prior higher education institution (HEI), former (type of) HEI, amount and quality of research experience, or composite scores). If one study addressed different methods or evaluative quality principles, that study was included in all respective categories. The number of papers cross-tabulated according to selection method and evaluative quality principle are presented in Table A2 in the Appendix. Table A3 in the Appendix shows the main characteristics of studies such as study design, country, field of study, and so forth. Table A3 also includes the summary of the relevant findings per study. The findings are synthesized below.

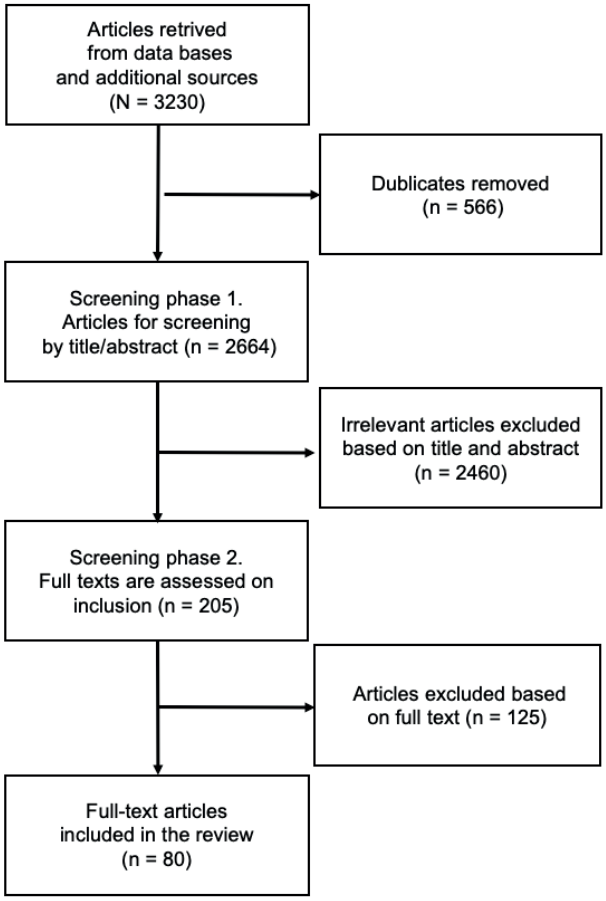


Figure 1| *Flowchart of Articles' Selection.*

2.3 Results

Table 2 in the Appendix shows the numbers of articles on each selection method and evaluative quality principle. We note the overall lack of research on the topics of reliability and cost-effectiveness. Therefore, the evidence below is presented mostly on validity, acceptability, and procedural issues. When studies on reliability or cost-effectiveness are available, they are reported in the respective selection methods' categories.

2.3.1 Prior Grades

2.3.1.1 Validity and Reliability of Prior Grades

The research focused on exploring the predictive validity of different aspects of grade point average (GPA), such us undergraduate GPA (UGPA), the first-year GPA, and the last-year GPA. Findings are presented in Table 2. Overall, it appears that UGPA is a valid predictor of student performance on introductory graduate courses (Moneta-Koehler et al., 2017; Park et al., 2018; Willcockson et al., 2009) and graduate GPA (GGPA; Bridgeman et al., 2009; Burton & Wang, 2005; Fu, 2012; Howell et al., 2014; Moneta-Koehler et al., 2017; Zimmermann et al., 2015). However, UGPA is not valid for predicting research productivity (defined as number of published papers, presentations, and obtained grants; Howell et al., 2014; Moneta-Koehler et al., 2017). UGPA is also not a valid predictor toward passing qualifying exams (Burmeister et al., 2014; Moneta-Koehler et al., 2017). There is mixed evidence on predictive validity of UGPA toward graduate degree completion (Cox et al., 2009; Dore, 2017; Moneta-Koehler et al., 2017; Wollast et al., 2018), time to graduate degree (Dabney, 2012; Howell et al., 2014; Moneta-Koehler et al., 2017), and faculty ratings (Howell et al., 2014; Moneta-Koehler et al., 2017). The directions of effects for these mixed findings are presented in Table 2.

Some single studies looked at UGPA in more detail. Namely, they disentangled UGPA on subsections such as first-year UGPA or last-year UGPA. A study that tried to predict graduate degree completion with first-year UGPA found no such relationship (DeClou, 2016). Two studies that explored the predictive validity of last-year UGPA found that last-year GPA is positively related to graduate degree completion (Dore, 2017) and GGPA (Zimmermann et al., 2017).

We found only one study that addressed the question of reliability estimates and it showed that the reliability estimates for undergraduate grade point average (UGPA) are relatively high, ranging between .89 and .92 (Westrick, 2017).

Table 2 | *Research Evidence on Validity of Prior Grades.*

Valid for the following dimensions of study success (References)	Exceptions or additional findings	Mixed /not sufficient evidence for the following dimensions of study success (References)	Not valid for the following dimensions of graduate success (References)
Undergraduate grade point average (UGPA)			
Graduate grade point average (GGPA) (Bridgeman et al., 2009; Burton & Wang, 2005; Fu, 2012; Howell et al., 2014; Moneta-Koehler et al., 2017; Zimmermann et al., 2015)	The overall good predictive power of UGPA differs per field of study. UGPA has a stronger predictive validity toward GGPA in chemistry departments than it has in biology departments (Burton & Wang, 2005). It also depends on how much narrowed the range UGPA in a study is (Burmeister et al., 2014).	**Graduate degree completion** *Positive relationship* (Moneta-Koehler et al., 2017; Wollast et al., 2018). *No relation* (Cox et al., 2009; Dore, 2017)	**Research productivity,** defined as number of published papers, presentations and obtained grants (Howell et al., 2014; Moneta-Koehler et al., 2017)
Performance on introductory courses (Moneta-Koehler et al., 2017; Park et al., 2018; Willcockson et al., 2009)		**Time to graduate degree** *Negative relationship* (Howell et al., 2014) *No relation*: (Dabney, 2012; Moneta-Koehler et al., 2017) **Faculty ratings** *Positive relationship*: ratings (Moneta-Koehler et al., 2017) *No relation*: (Howell et al., 2014)	**Passing the qualifying exam** (Burmeister et al., 2014; Moneta-Koehler et al., 2017)
The first-year undergraduate GPA			
		Graduate degree completion *No relationship* (DeClou, 2016)	
The last-year undergraduate GPA			
		GGPA *Positive relationship, the strongest predictor* (Dore, 2017) **Graduate degree completion** *Positive relationship, the strongest predictor* (Zimmermann et al. 2018)	

2.3.1.2 Procedural Issues of Prior Grades

The procedural issues with using prior grades for admissions decisions are grade inflation and differences in grading standards. Grade inflation is a practice of awarding higher grades than previously assigned for given levels of achievement (Merriam-Webster dictionary, n.d.). For example, teachers giving higher grades for positive student ratings (European Grade Conversion System [EGRACONS], 2020). In her observational study of top graduate research programs, Posselt (2014) indicated that grade inflation is a widespread phenomenon in highly selective universities. In such universities, students from underrepresented backgrounds are extremely lacking, therefore, setting a grade-threshold on a high level disproportionately excluded these students (Posselt, 2014).

Another procedural issue—differences in grading standards—relates to the fact that one grade obtained at different institutions might reflect a different level of academic qualification. Grade conversion and grade distribution tables, which are developed to tackle these issues, are not without limitations. They can often be crude, and this can affect both selection decisions and research done on grades as predictors of graduate study success (see, e.g., Zimmermann et al., 2017).

2.3.1.3 Acceptability of Prior Grades

Prior grades are a widely accepted selective admissions method (Boyette-Davis, 2018; Kurysheva et al., 2019; MasterMind Europe, 2017). The largest weight in admissions decisions is given to grades on undergraduate courses that are closest in terms of content to the courses of a graduate program (Chari & Potvin, 2019). When explaining what the reasons are behind high acceptability of grades and even overestimation of their importance in graduate admissions by admissions committees, Posselt (2014) states that high conventual achievements, such as grades, are consistent with the identity of an elite intellectual community, which admissions committee members, implicitly or explicitly, refer themselves.

2.3.2 Standardized Testing of Academic Abilities

2.3.2.1 Validity of Standardized Admissions Tests of Academic Abilities

Among different standardized admissions tests, the ones which are typically required for selective admissions to graduate programs in STEM disciplines are the Graduate Record Examinations (GRE) General and GRE Subject. All but one study, which addressed validity of standardized tests, referred to these two GRE tests. The only exception was the standardized test EXANI-III, which is used in Mexico.

Validity of graduate standardized admissions tests has been a controversial topic in research, with some studies providing evidence for their weak-to-moderate predictive power toward graduate study success and others indicating the absence of predictive power (see Table 3). From Table 3, we can infer that the standardized test most often

examined is the GRE General. The GRE General is a positive predictor of first-year GGPA, GGPA, and faculty ratings, which is in line with the existing reviews and meta-analyses (Kuncel et al., 2010; Kuncel & Hezlett, 2007b, 2010). From the majority of primary studies, it appears that the GRE General does not predict graduate degree completion and research productivity defined as the number of publications. The meta-analyses on the topic, however, found that after meta-analytical corrections for statistical artifacts in primary studies were applied (such as a correction for the restriction of range of a predictor), these two relationships (1) between the GRE General and degree completion and (2) between the GRE General and research productivity, although weak, were detected (Kuncel & Hezlett, 2007a, 2007b). Finally, there was mixed or limited evidence for GRE General efficiency in prediction of time to graduate degree, performance on core program courses, qualifying exam, rate of progress, and thesis performance (see Table 3 for details).

There is an indication that another standardized test, the GRE Subject in Physics, is predictive for faculty ratings, while its predictive value for graduate degree completion remains unclear. Two meta-analyses also found that the GRE Subject is a meaningful predictor of graduate study success (Kuncel et al., 2010; Kuncel & Hezlett, 2007b).

2.3.2.2 Procedural Issues of Standardized Admissions Tests of Academic Abilities

The primary studies showed a possibility of adverse impact of the GRE on underrepresented groups (including ethnic minorities and females in STEM) as well as item position effects. The former can be mitigated by applying a systematic and holistic approach in reviewing admissions files (Bleske-Rechek & Browne, 2014; Murphy, 2009; Posselt, 2014; Wilson et al., 2018; Wilson et al., 2019), while the latter can be mitigated by allowing proper time limits for taking the test (Davey & Lee, 2011). However, the reviews and meta-analyses on procedural issues refuted several common beliefs regarding standardized tests, such as the coaching effects, lack of predictive independence from socioeconomic status (SES), and bias in testing. The coaching effects were shown to be modest with one quarter of a standard deviation improvement in test performance (Hausknecht et al., 2007; Kuncel & Hezlett, 2007). Such an improvement refers primarily toward the GRE Analytical Writing section (GRE-A) (Powers, 2017). GRE Verbal Reasoning (GRE-V) and GRE Quantitative Reasoning (GRE-Q) were prone to coaching to a negligible extent in contrast to claims of commercial organizations that prepare test takers for standardized tests (Powers, 2017). Lack of predictive independence from SES was contested by demonstrating that even after controlling for SES, standardized test scores remained predictive of study success (Camara et al., 2013; Kuncel & Hezlett, 2010). According to one meta-analysis and one review, bias in graduate testing is a myth, as, according to their findings, standardized tests appeared to predict graduate study success of both females and males equally (Fischer et al., 2013; Kuncel & Hezlett, 2007b) as well as ethnic groups (Kuncel & Hezlett, 2007b).

Table 3 | *Research Evidence on Validity of Standardized Tests of Academic Abilities.*

Valid for the following dimensions of study success (References)	Exceptions or additional findings	Mixed /not sufficient evidence for the following dimensions of study success (References)	Not valid for the following dimensions of study success (References)	Exceptions or additional findings
		GRE General		
First-year graduate GPA (Bridgeman et al., 2009; Burmeister et al., 2014; Fu, 2012; Moneta-Koehler et al., 2017)		**Time to graduate degree** *No relationship:* (Dore, 2017; Hall et al., 2017; Moneta-Koehler et al., 2017; Petersen et al., 2018; Sealy et al., 2019) *Positive relationship*: the higher the GRE scores, the longer the time (Howell et al., 2014; Lorden et al., 2011) *Negative relationship*: the higher the GRE scores, the shorter the time (Frasier, 2013), also depending on the field of study (Miller, 2013)	**Graduate degree completion** (Cox et al., 2009; Dore, 2017; Lorden et al., 2011; Lott et al., 2009; Miller et al., 2019; Moneta-Koehler et al., 2017; Petersen et al., 2018)	One study (Perez, 2011) is an exception: GRE-V was found to be predictive for degree completion of both masters' and doctoral students, and GRE-Q was predictive only for degree completion of masters' students. Interestingly, when a relative GRE score is considered instead of its absolute score (i.e., compared to student's peers in a program), one study has found GRE General to be efficient in distinguishing students with lower and higher odd of attrition in one of the studies (Lott et al. 2009).
GGPA (Burton & Wang, 2005; Howell et al., 2014; Klieger et al., 2014; Moneta-Koehler et al., 2017; Perez, 2011; Zimmermann et al., 2018)	One study did not find a correlation of any GRE General sections to GGPA, but it used self-reported data on GGPA, and the sample lacked variability in GRE (Sanford, 2009).	**Performance on core program courses** The findings are ambiguous: one study found that the correlation is positive (Burmeister et al., 2014), another one showed that while the GRE General contribution existed in univariate analysis, it did not hold in the adjusted model (Park et al., 2018). There is another study showing that only GRE-V, but not GRE-Q was positively related to a core course performance (Willcockson et al., 2009).	**Research productivity** defined as number of publications (Hall et al., 2017; Moneta-Koehler et al., 2017; Sanford, 2009; Sealy et al., 2019)	One study is an exception, where GRE-Q was the only and very weak predictor of number of publications, explaining 5% of variance in the outcome (Howell et al., 2014)

Faculty ratings (Burmeister et al., 2014; Burton & Wang, 2005; Howell et al., 2014; Moneta-Koehler et al., 2017)	An exception is a study that found the negative relationship between the GRE General and faculty rankings: better ranked PhD students tended to have lower GRE scores from their faculty mentors (Sealy et al., 2019). Though the authors state that they had a wide span of the GRE scores in their data, the sample size in this study was extremely low: 28 students and part of the analysis was conducted on the lower and upper quartiles of the GRE scores, lowering the sample size even more.	**Qualifying exam** *No relationship* (Moneta-Koehler et al., 2017) *Positive relationship* (Burmeister et al., 2014)	**Conference presentations** (Moneta-Koehler et al., 2017)	
		Rate of progress GRE-Q is a weak predictor of rate of progress (Zimmermann et al., 2018)	**Obtaining grants, fellowships, or awards** (Moneta-Koehler et al., 2017; Sealy et al., 2019)	
		Thesis performance *No relationship*: Zimmermann et al., 2018		
Graduate Record Examinations Subject (Physics)				
Faculty ratings (Burmeister et al., 2014)			**Graduate degree completion** (Miller et al., 2019)	The study of Miller et al. (2019) was heavily criticized for several issues in research design and statistical approach (Weissman, 2020), thus suggesting that the finding must be regarded with a lot of caution.
Other standardized tests (EXANI-III in Mexico)				
			Graduate degree completion (Álvarez-Montero et al., 2014)	

The authors of these studies also indicated that the differences in performance between different groups might reflect societal problems such as lack of family, social, environmental, peer, and financial support. They state that standardized tests simply expose the preexisting differences created by the above-mentioned societal problems (Camara et al., 2013; Kuncel & Hezlett, 2010). One review study brought attention to the problem of a negative effect of stereotype threat on standardized test performance (Garces, 2014): Test takers, who believe that their nonoptimal performance on standardized tests might confirm the stereotypes of their minority group's intellectual capacity, might perform worse because of that self-fulfilling prophecy.

2.3.2.3 Acceptability of Standardized Admissions Tests of Academic Abilities

Acceptability by Admissions Committees. In the US context, admissions committees—especially for research programs—actively use the GRE General and consider it to be a valuable contributor for their admissions decisions (Boyette-Davis, 2018; Chari & Potvin, 2019; Rock & Adler, 2014). Out of the three sections, GRE-V and GRE-Q are used most, while GRE-A is considered the least often (only around 35% of surveyed programs; Briihl & Wasieleski, 2007). When it comes to positioning GRE as a selection method, the GRE appeared less important than, for example, previous research experience, UGPA, and certain personal characteristics (e.g., critical thinking, work ethics; Boyette-Davis, 2011). However, the GRE had more weight in selection decisions for doctoral programs than for masters' programs (Chari & Potvin, 2019).

A survey among masters' programs in Europe showed that the results of standardized admissions tests are rarely used for elimination purposes (only around five percent of the masters' programs admitted such a practice), but higher scores, if present, do provide an advantage to students in one fourth of the programs (MasterMind Europe, 2017). However, Europe has seen a steady increase in GRE test takers (e.g., it increased from 12,243 in 2004 to 29,211 in 2013) since the introduction of the Bologna Process and the increasing internationalization of European graduate education (Payne, 2015). Test takers aiming to study STEM disciplines represented the largest group among all European GRE test takers (Payne, 2015).

Acceptability by Applicants. Applicants viewed the GRE as less important in graduate admissions than UGPA, recommendation letters, and work experience (Cline & Powers, 2014). Applicants coming from racial minority groups had more negative feelings about the GRE than white test takers (Cline & Powers, 2014). International students felt that the GRE is culturally biased (Mupinga & Mupinga, 2005). Applicants perceived publishing prompts from GRE-A positively (Powers, 2005) and desired to get additional information about their writing skills beyond their GRE-A score (Attali & Sinharay, 2015).

2.3.2.4 Cost-Effectiveness of Standardized Admissions Tests of Academic Abilities

One study looked at this evaluative quality principle. In their study, Klieger et al. (2014) provided an example of calculation of the benefits for one US doctoral program. They estimated the financial benefits of using the GRE for admissions and funding decisions as considerable, but obviously, the exact numbers will depend on a specific program and a number of GRE sections used for admissions decisions.

2.3.3 Letters of Recommendation (LoRs)

2.3.3.1 Validity and Reliability of Letters of Recommendation

The only primary study which examined predictive validity of LoRs for STEM disciplines (namely the biomedical sciences) found that the scores on LoRs did not predict time to degree, but they were the most powerful predictor of first-author student publications (Hall et al., 2017). The review of Kuncel et al. (2014) showed that LoRs do not deliver incremental validity over standardized admissions tests and UGPA toward GGPA and faculty ratings but do deliver small incremental validity in prediction of degree completion (an outcome usually difficult to predict using other measures). The review of Megginson (2009) showed that narrative LoRs have minimal reliability and are prone to subjective interpretations.

2.3.3.2 Procedural Issues of Letters of Recommendation

Three primary studies that explored biases in narrative LoRs at the graduate level found evidence of gender and race biases (Biernat & Eidelman, 2007; Morgan et al., 2013), bias arising from tone of LoRs, and bias arising from admissions committees' members being (un)familiar with the LoR writer (Posselt, 2018). In addition, the study of Morgan et al. (2013) found that allowing the applicants' names to be visible on LoRs negatively influenced admissions committees' evaluations of underrepresented minority groups. Requiring admissions committees to elaborate on their evaluations of narrative LoRs reduces biases (Morgan et al., 2013).

2.3.3.3 Acceptability of Letters of Recommendation

Two primary studies explored the acceptability of LoRs. One study showed that LoRs are the second most valued selection method in admissions to doctoral programs in the US context, because they shed light on applicants' personal characteristics (Boyette-Davis, 2018). However, another study in the European context did not find that LoRs are given weight by admissions committees when they decide to reject or admit a student to a master's program (MasterMind Europe, 2017). In the latter study, more than a half (58.3%) of surveyed applicants reported that they had to provide a LoR within their application file.

2.3.4 Interviews

2.3.4.1 Validity of Interviews

Evidence on validity of interviews in STEM graduate programs is limited to two studies. One focused on traditional interviews and the other on the highly structured and formalized form of interviews: multiple mini-interviews (MMIs). Traditional interviews do not allow to distinguish between most and least productive graduate students (in terms of their time to degree and number of first-author papers; Hall et al., 2017). However, MMIs allow to predict planning-related problematic study behavior (oude Egbrink & Schuwirth, 2016).

2.3.4.2 Procedural Issues of Interviews

No study addressed the procedural issues of interviews specifically in graduate admissions.

2.3.4.3 Acceptability of Interviews

A survey among European masters' programs demonstrated that interviews are used in 22.6% of English-taught masters' programs across Europe (MasterMind Europe, 2017). Although it is not a widely used selection method, it is valued and regarded as a good practice by admissions committees. Additionally, members of admissions committees reported that a poor interview is a reason for rejection in less than 5% of all cases. No studies were conducted on how favorable interviews are perceived by applicants to graduate programs.

2.3.5 Personal Statements (Motivation Letters)

2.3.5.1 Validity of Personal Statements

A meta-analysis on predictive validity of personal statements showed that they were weak predictors of grades and faculty ratings and when considered together with the UGPA and standardized admissions tests, they provided no incremental validity (Murphy et al., 2009). An interesting case is a primary study that developed a five-point rubric for measuring intentionality aspects in personal statements, such as motivation, beliefs, goal-setting, and self-efficacy (Jones, 2013). The construct validity of this rubric was shown to be sufficient, but the predictive validity has not been explored yet (Jones, 2013).

2.3.5.2 Procedural Issues of Personal Statements

No studies examined procedural issues of personal statements specifically in graduate admissions.

2.3.5.3 Acceptability of Personal Statements

Personal statements are used frequently (Kurysheva et al., 2019; MasterMind Europe, 2017), and are required from international applicants almost twice as often as from internal applicants (i.e., those, who obtained a bachelor's degree at the same institution; MasterMind Europe, 2017). Personal statements are used to assess students' motivation but also to make inferences about personal qualities, previous academic background, and cognitive ability (Kurysheva et al., 2019). They also help provide information on whether a student's background will contribute to the diversity of the student body (Posselt, 2014). In most cases, personal statements did not serve as a reason for failure in the admissions process, according to members of admissions committees (MasterMind Europe, 2017).

2.3.6 Intelligence Assessments

2.3.6.1 Validity of Intelligence Assessments

Intelligence assessments are significantly correlated with academic performance (defined as grades, results of educational tests, and procedural and declarative knowledge; Poropat, 2009; Schneider & Preckel, 2017).

2.3.6.2 Procedural Issues of Intelligence Assessments

Practical utility of intelligence as a predictor of study success is usually reduced because it overlaps significantly with measures of prior performance (e.g., grades; Poropat, 2009).

2.3.6.3 Acceptability of Intelligence Assessments

In a cross-sectional study on the samples of students in the life sciences and natural sciences, it was shown that admissions criteria related to intelligence play a moderately important role in admissions decisions along with several other admissions criteria (Kurysheva et al., 2019). However, those admissions committees participating in the study did not apply specific intelligence assessments in their programs; the inferences on student intelligence were made from other selection methods rather than specific intelligence testing (Kurysheva et al., 2019).

2.3.7 Personality Assessments

2.3.7.1 Validity of Personality Assessments

The most common personality assessment is based on the five-factor model named the "Big Five". It distinguishes five primary factors of personality: extraversion, agreeableness, openness to experience, conscientiousness, and neuroticism (Goldberg, 1993). One of the most stable findings both from individual and meta-analytical studies

is that one of the Big Five personality traits (conscientiousness) is a medium-to-large predictor of study success (Butter & Born, 2012; Poropat, 2009; Schneider & Preckel, 2017; Trapmann et al., 2007; Walsh, 2020). When it came to other personal traits, measured by the Big Five, it appears that extraversion and neuroticism have no significant relation to study success, while there are mixed findings regarding agreeableness and openness to experience (Poropat, 2009; Trapmann et al., 2007).

Other personal traits, not explicitly included in the Big Five, were also examined: grit (defined as determination to achieve long-term goals), emotional intelligence, need for cognition (defined as an inclination to value activities that include effortful cognition), and conscientiousness related to time management, so-called ecological conscientiousness. It was found that grit does not explain additional variance in study success beyond conscientiousness (Walsh, 2020), while emotional intelligence and need for cognition do have a weak-to-moderate effect (Schneider & Preckel, 2017). One study developed and validated a scale for ecological conscientiousness and found that it is valid beyond the conventional Big Five in predicting Ph.D. performance criteria such as research progress, meeting deadlines, and probability to obtain a Ph.D. degree on time (Butter & Born, 2012).

2.3.7.2 Procedural Issues of Personality Assessments

Two procedural issues of personality assessments are referred to in the context of graduate admissions: applicant faking and their coachability (Kyllonen et al., 2005). They arise from the fact that personality assessments are typically based on self-reports.

2.3.7.3 Acceptability of Personality Assessments

While graduate admissions committees regard personality assessment important to consider in principle (Kyllonen et al., 2005), they do not report to use them extensively (Boyette-Davis, 2018; Kurysheva et al., 2019; MasterMind Europe, 2017).

2.3.8 Language Proficiency Assessments

2.3.8.1 Validity of Language Proficiency

The available evidence on validity of different language assessments toward different dimensions of study success is presented in Table 4. From research on the Test of English as a Foreign Language (TOEFL), it appears that out of four languages skills assessed by the test (writing, listening, reading, and speaking), the writing section had the greatest potential (Fu, 2012) and is advised to be considered during selective admissions (Bridgeman, 2016).

Table 4 | *Research Evidence on Validity of Language Assessments.*

Valid for the following dimensions of study success (References)	Exceptions or additional findings	Mixed /not sufficient evidence for the following dimensions of study success (References)	Not valid for the following dimensions of study success (References)
		Test of English as a Foreign Language (TOEFL)	
Graduate GPA *Small positive relationship* (Abunawas, 2014; Arcuino, 2013; Cho & Bridgeman, 2012; Fu, 2012; Wongtrirat, 2010; Zimmermann et al., 2018)	One study is an exception, where no relation was found (Sanford, 2009)	**First-year GPA** *The correlation exists* (Burmeister et al. 2014). Some studies find the incremental value of TOEFL (Cho & Bridgeman, 2012; Zimmermann et al., 2018), others do not (Fu, 2012).	**Graduate degree completion** (Sanford, 2009; Wongtrirat, 2010)
		Course average; Faculty ratings *Positive relationship* (Burmeister et al., 2014)	**Time to graduate degree** (Sanford, 2009; Wongtrirat, 2010)
		International English Language Testing System (IELTS)	
		Graduate GPA *No relationship* (Arcuino, 2013)	
		The Computerized Enhanced ESL Placement Test (CEEPT)	
		First semester academic performance *Mixed findings* (Lee & Greene, 2007)	
		A scale that takes into account the nature of the previous language use	
		Completing a PhD degree in a foreign HEI *A positive relationship* (Mathews, 2007)	

2.3.8.2 Procedural Issues of Language Proficiency Assessments

No studies were detected that examined procedural issues of language proficiency assessments, specifically for graduate admissions.

2.3.8.3 Acceptability of Language Proficiency Assessments

A study on the European context found that English language assessments were required mostly from foreign applicants to masters' programs, although internal applicants are sometimes expected to submit them as well (MasterMind Europe, 2017). Perceived importance of language proficiency by faculty members depended on a discipline: In humanities, for example, the importance is higher than in science disciplines (Lee & Greene, 2007). Admissions committees usually limit the usage of language proficiencies assessments by checking whether the institutional cut-off score was met. Faculty members often expressed dissatisfaction with the language proficiency of admitted students because some of them think the cut-offs reflect not adequate but only minimal required language proficiency (Ginther & Elder, 2014). In line with that, test takers do not seem to perceive TOEFL scores as a good indication of one's language abilities (Mathews, 2007).

2.3.9 Prior Research Experience

2.3.9.1 Validity of Prior Research Experience

The results of studies on validity of prior research experience are mixed. Three studies found predictive power of prior research experience for research skills performance (Gilmore et al., 2015), doctoral degree completion (Cox et al., 2009), and faculty ratings (Weiner, 2014). However, two other studies did not detect a relation between prior research experience and performance on an introductory graduate biomedical course (Park et al., 2018). Nor did they detect graduate student productivity or time to degree (Hall et al., 2017).

2.3.9.2 Procedural Issues of Prior Research Experience

No studies examined procedural issues of prior research experience specifically in graduate admissions.

2.3.9.3 Acceptability of Prior Research Experience

It appears that prior research experience is a valued component in graduate admissions (Boyette-Davis, 2018; Chari & Potvin, 2019; Kurysheva et al., 2019). However, the extent of its importance depends on whether it is applied to a master's or a doctoral program level (Chari & Potvin, 2019). The extent of its importance also depends on what aspects of prior research experience are available for review. For example, simply having a basic level of research experience is significantly more important than having publications or conference participation records (Boyette-Davis, 2018).

2.3.10 Various Graduate Selection Methods

In this category of various graduate selection methods, the selection methods were collected that did not fall in previously reviewed categories. Namely, undergraduate institution selectivity, type of prior degree (bachelor's or master's), composite score of different selection methods, specific assessment instruments, same field or same university as the one during the prior degree, rate of progress, and duration of prior studies.

2.3.10.1 Validity of Various Graduate Selection Methods

Undergraduate institution selectivity appears to have a positive relation to performance during the first semester of graduate studies (Moneta-Koehler et al. 2017, Park et al. 2018). Having a prior graduate degree increases the chances of graduate study success (Willcockson et al., 2009). The last three sub tables of Table A.3 in the Appendix (A3.24-A3.26) provide details into the findings of single studies on validation of all selection methods, which fell in this category.

2.3.10.2 Procedural Issues of Various Graduate Selection Methods

Due to the scarcity of validation studies of the selection methods in this category, the procedural issues remain underexamined. One study addressed academic pedigree as a procedural issue of undergraduate institution selectivity (Posselt, 2018). Academic pedigree is the belief that higher rank of prior HEI signifies stronger student performance potential. In case of academic pedigree, the grades might be interpreted within the context of how rigorous the student's curriculum was at a prior HEI. However, it appears that the selectivity and reputation of prior HEI are not clearly stated but somewhat hidden selection methods (Posselt, 2018). Posselt (2018) underscored that "privileging elite academic pedigrees in graduate admissions preserves racial and socioeconomic inequities that many institutions say they wish to reduce" (p. 497).

2.3.10.3 Acceptability of Various Graduate Selection Methods

Certain undergraduate courses, type of prior academic background, and type of prior education institution are considered decisive factors in admissions by graduate admissions committees (Chari & Potvin, 2019; Kurysheva et al., 2019). At the same time, the whole range of various selection methods such as extracurricular activities, teaching experience, quantitative skills, work experience, curriculum vitae (CV), photographs, essays, time management skills, understanding social relevance of research, and evidence of integrity, even if required, were not given substantial weight in selection decisions (Boyette-Davis, 2018; Kurysheva et al., 2019; MasterMind Europe, 2017). Assessment methods that consist of different scales that range from scientific knowledge to motivation are generally not opposed by applicants (van Os, 2007).

2.4 Discussion

This study, which focuses on the available research between 2005–2020, is the first review on both cognitive and noncognitive selection methods in graduate education and focuses on STEM disciplines. Studies dedicated to reliability and cost-effectiveness of graduate selection methods were rarely conducted during the examined time span. Therefore, the review's focus was on integrating research evidence on the three evaluative quality principles of predictive validity, acceptability, and procedural issues.

2.4.1 Summary: Key Findings

It was shown that the predictive validity of applied selection methods varies substantially and depends on several factors (e.g., field of study under examination or which other selection methods are included in the research model of a study). The procedural issues of selection methods were also reviewed: admissions biases, faking, coaching effects, item position effects, test preparation, and stereotype threat. As for acceptability by admissions committees, an interesting finding was that some invalid selection methods are widely accepted by admissions committees (e.g., traditional interviews), while a similar method with a more structured format and with preliminary indications for validity (e.g., MMI) do not appear to be widespread in STEM admissions.

Figure 2 provides a visualization of the selection methods located according to the extent of their predictive validity on the horizontal axis and acceptability by admissions committees on the vertical axis. The dimensions of acceptability by applicants or procedural issues are not depicted because this would require a third and fourth dimensions which would make the figure more difficult to interpret. The names of selection methods are typed in a larger font. The dimensions of graduate study success (which predictive validity was evaluated) are typed under the names of selection methods in a smaller font. Three colors are used: red, light green, and dark green. They denote negligible, small-to-medium, and medium-to-strong predictive validity toward the examined dimensions of graduate study success, respectively. Figure 2 is given for illustrative purposes only and it does not purport reflecting the details that are presented in the text of this review.

When summarizing the findings on predictive validity, we can infer from this review that prior grades (including UGPA), GRE General, intelligence assessments, and the personality trait conscientiousness are medium-to-strong predictors of several graduate study success dimensions. Letters of recommendation, tests on language proficiency, and personality aspects such as emotional intelligence and need for cognition are also valid predictors but to a lesser extent. There are some indications that such selection methods as undergraduate research experience and MMIs are also sensible selection instruments in the context of graduate STEM disciplines, but the research on them is scarce. The selection methods in graduate admissions with lack of predictive validity

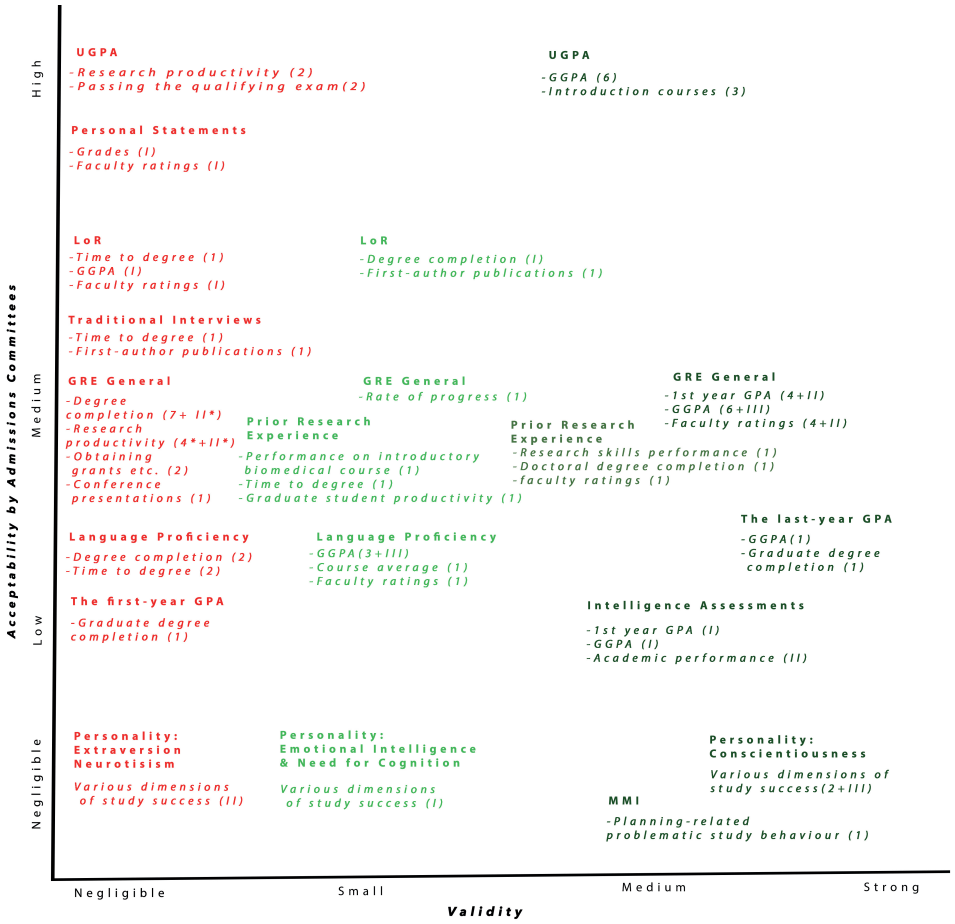


Figure 2 | *The Summary of the Findings on Two Evaluative Quality Principles: Validity and Acceptability by Admissions Committees.*

Note. X-axis and Y-axis represent two evaluative quality principles: (1) validity and (2) acceptability by admissions committees, respectively. The location of selection criteria (**in bold**) and the respective dimensions of study success (*in italics*) are approximations based on the findings of the review. Arabic and Roman numbers are indicated in parentheses. The Arabic numbers next to each examined dimension of study success stand for the number of primary studies on which the finding is based. The Roman numbers stand for the number of secondary studies (such as reviews and meta-analyses). The colors refer to the X-axis: Red is used for selection methods with respective dimensions of study success toward which they are invalid. Light green and dark green are used for selection methods that have small-to-medium and medium-to-strong extent of predictive validity, respectively, toward examined dimensions of study success.

Disclaimer. Because calculation of the exact validity and acceptability estimates was not the goal of this review, the reader should account for the fact that the location of a selection method is an approximation. Next, because this figure presents the generalized findings on predictive validity and their acceptability by stakeholders, it should always be interpreted in the context of findings of procedural issues of these methods, their acceptability by students, and cost-effectiveness. For detailed findings on each selection method, the reader is referred to Tables 1, 2, 3, Table A.3 in the Appendix, and the texts of the studies.

were detected by this review as well: personal statements, traditional interviews, and two personal traits (extraversion and neuroticism). Certain methods (e.g., the GRE General and UGPA) were examined extensively and would appear valid toward certain dimensions of study success (e.g., GGPA), but not the others (e.g., research productivity).

As seen in Figure 2, the acceptability of graduate selection methods by graduate admissions committees is not always in line with the extent of their validity. For example, personal statements appear to have negligible validity, especially in the presence of other selection methods but are still widely used. From the review, it is also evident that no selection method comes without procedural issues. While for some of the methods, the procedural issues constitute a prominent research debate (e.g., a debate on biases involved in implementation of the GRE), the procedural issues of others have not been adequately addressed (e.g., imperfections of grade conversion).

2.4.2 Some Evidence from outside of STEM Graduate Admissions

It is important to note that there is profound research on procedural issues and acceptability of selection methods outside of graduate admissions, namely in undergraduate admissions and personnel selection. They were not included in results because they did not fulfill inclusion criteria for this review. However, they are worth mentioning here in the discussion section because it is unlikely that the procedural issues of the same selection method such as biases, faking, or coaching would be heavily determined by the education level. The following two subsections (procedural issues and acceptability) will, therefore, be dedicated to the outline of those procedural issues and acceptability of some selection methods that received little attention in graduate admissions but were investigated in undergraduate admissions and personnel selection.

2.4.2.1 Procedural Issues

Procedural Issues of (Traditional) Interviews. The current review did not detect studies on procedural issues of interviews in graduate admissions. However, the findings from undergraduate and personnel are as follows. The first procedural issue is susceptibility of interviews to biases toward gender, disability status, and ethnicity. Biases during interviews might come into play at different moments starting from so-called rapport building (a "small chat" aimed at helping applicants to feel comfortable), through the interview itself, and during the evaluation stage after the interview has ended (Levashina et al., 2014). Reducing bias and increasing validity and reliability of interviews is possible through introducing structure and different formats of interview: For example, phone or video interviews are more adaptable for structuring than face-to-face interviews (Levashina et al., 2014). The second procedural issue is susceptibility of interviews to subjective interpretations of student "soft variables", such as motivation. A study on a sample of students in a selective college in the Netherlands demonstrated that scores on interviews contribute little to prediction of study success but create risk

of subjective interpretations. For example, many of the students whom the interviewers indicated were at risk of expulsion finished their first year successfully (Reumer & van der Wende, 2010). The authors note that "interviews provide extra guidance to both the student and the institution as to whether the student is choosing the right study program (and not so much as whether he is able to complete it successfully)" (Reumer & van der Wende, 2010, p. 20). The third procedural issue of interviews is faking by applicants, defined as "the conscious distortions of answers to the interview questions in order to obtain a better score on the interview and/or otherwise create favorable perceptions" (Levashina & Campion, 2007, p. 1639). Among undergraduate job applicants, the estimates of faking, understood in the above-defined broad sense, are as high as 90%, and the estimates of faking that is closer to lying range from 28% to 75% (Levashina & Campion, 2007).

Finally, in her ethnographical study on graduate admissions[10], Posselt (2016) points out several other issues with admissions interviews. She draws attention to impression management strategy used by some applicants (e.g., constant smiling), which contributes to admissions committees' perception of these applicants as "glowing" and having "a very nice personality" (Posselt, 2016, p. 144). She also points out the issue of admissions committees' distrust toward language skills of certain groups of international applicants (Posselt, 2016). Finally, interviews provoke a broader actual evaluation of applicants than is formally communicated (e.g., proclaiming the aim of the interview to be only an additional language check, while pursuing the aim to get an impression of personal qualities of an applicant).

Procedural Issues of Personal Statements. In the literature outside of graduate selection, namely in the medical education programs, the biases of gender, age, socioeconomic class, country of origin, and ethnicity were shown to be present in admissions committees' evaluations of personal statements (for the description, see the review of Kuncel et al., 2020).

Procedural Issues of Personality Assessments. Similar to findings in graduate admissions, researchers who conducted studies in undergraduate and personnel selection show that the major procedural issue appears to be faking (Birkeland et al., 2006; König et al., 2017; Pavlov et al., 2019). The extent of faking depends on personality dimension under examination, type of test, aimed position (Birkeland et al., 2006), and situation stakes (Pavlov et al., 2019). However, there are approaches where supervisors of students are asked to report on their personality, and while the supervisors also tend to fake when reporting on the personality of their students, the extent of their faking is smaller (König et al., 2017).

10 It was not included in the "Results" section of this review as it is an authored and not an edited book.

2.4.2.2 Acceptability

In personnel selection, a review was conducted on how favorable different selection methods are rated by job applicants. From the review, it appears that the most preferred methods are work sample and interviews; overall favorably evaluated selection methods are resumes, cognitive tests, references, and personality assessments. The least preferred are honesty tests, personal contacts, and graphology (Anderson et al., 2010). Each selection method was assessed on several acceptability scales. For example, perceived scientific validity of LoRs is low, but their interpersonal warmth is high. In contrast to LoRs, intelligence assessments are perceived high on scientific validity and respectful of privacy but low on interpersonal warmth (Anderson et al., 2010). Interestingly, when it comes to structure of interviews, both applicants and interviewers perceive structured interviews less positively than unstructured interviews (Levashina et al., 2014). Similar to interviews, applicants perceive personality assessments favorably, especially the dimension "opportunity to perform" (Anderson et al., 2010).

2.4.3 Graduate Selection Methods as a Distinct Area for Research

This review maps research evidence on selection methods used specifically at the graduate level. Several selection instruments that are used in admissions to professional schools such as medical school (e.g., situational judgment tests, MMIs, and selection centers) are not used in graduate STEM admissions. What are the potential reasons for this difference? The most obvious difference is that admissions to professional schools are directed toward detecting certain skills and traits of applicants to predict key competencies which are different from those of STEM researchers. The frameworks have been developed that define key competencies in medical profession (e.g., the Canadian Medical Education Directives for Specialists). They specify the knowledge, skills, abilities, and other characteristics (KSCAOs), related to competent performance within certain healthcare professions (for examples, see Kerrin et al., 2018). Like medical education, graduate STEM education is also confronted with the question of which KSCAOs define an engineer or a researcher in STEM fields. A more general question would be even broader: whether a person is a researcher or a professional or not—and if not, why not? Does this have to do with academic freedom of researchers (Vrielink et al., 2011) and their roles as producers of critical knowledge, contributors to expansive learning, and organizers of a space for dialogue (Miettinen, 2004)? Do the existing selection instruments reviewed in this study adequately capture prerequisites for competent performance on researchers' roles? Are there any other selection methods that have potential to do this better? This review might, therefore, be regarded only as one of the first steps toward getting closer to answering such questions.

2.4.4 Strengths and Limitations

2.4.4.1 Strengths

The main strength of this review is that it synthesized high-quality research evidence across four evaluative quality principles of both cognitive and noncognitive selection methods. No such synthesis has been conducted in the field of graduate admissions in general, and in the STEM fields specifically. Another strength of this review is that based on the findings of both primary and secondary (reviews, meta-analyses) studies, this review compared them wherever possible. This allowed the detection of possible discrepancies such as range restriction in the primary studies. All things considered, this review provides a compilation of state-of-the-art research on graduate selective admissions in STEM fields of study. It has also clearly shown that the value of a selection method varies over different dimensions of study success.

2.4.4.2 Limitations

Drawing conclusions from a large number of papers inevitably brings a risk of losing the nuances of each study (see Table A3 in the Appendix for more details). It also means that the samples of studies on predictive validity of graduate selection methods in several instances included not only STEM students but also students from other disciplines. Even if the strength of the relationship between a selection method and various dimensions of graduate study success is diluted by inclusion of students from other disciplines, it is unlikely that the direction of relationship would be the opposite. From this, however, an advantage appeared that the findings of this review to a certain extent are generalizable to other academic disciplines within graduate levels of education.

Furthermore, our inferences on the effects sizes (negligible, small, medium, and strong effect sizes) were based on the interpretations of the studies' authors. To refine the estimations of the effect sizes, the meta-analyses on reviewed selection methods would be required. Such goals were outside the scope of this review, however, the indications that this review provides are robust enough to answer the main question on whether a selection method is valid in principle.

Finally, the reviewed literature on acceptability of selection methods often contained evidence from admissions committees' self-reports. Their reports could have been (un)consciously biased to a certain extent if they did not want to report, for example, the usage of invalid yet favored selection methods. Therefore, the observational ethnographic studies, like the one of Posselt (2016), gain special importance in this area of research: The observation might be a more appropriate method to detect "hidden" selection criteria and group dynamics within an admissions committee because these concealed processes are influential toward admissions decisions.

2.4.5 Implications for Theory, Practice, and Society

2.4.5.1 Implications for Theory

In most of the primary studies reviewed, the regression approach was used. While it is a widely accepted type of analysis in this field, it is limited because the findings on amount of explained variance are usually hard to interpret. Moreover, the findings based on the regression approach do not allow one to set the cut-off scores. Future research would benefit from applying other methodologies. For example, Bridgeman et al. (2009) offer a method that divides students within a department into quartiles based on a selection method of interest and a dimension of study success. The methodology that allows (under certain conditions) the establishment of cut-off scores for selective admissions methods is the Signal Detection Theory (van Ooijen-van der Linden, 2017). Finally, future research approaches toward selection methods should account for a multilevel and dynamic nature of student selection (Patterson et al., 2018) as well as the importance of other evaluative quality principles of selection methods not addressed in this review such as practicality/administrative convenience, ease of interpretation, and so forth (see for the full list Patterson & Ferguson, 2010).

2.4.5.2 Practical Implications

Research evidence on selection methods has advanced significantly in recent years. In some national and institutional contexts, the research findings are actively being translated into practice (e.g., Council of Graduate Schools, 2021). However, along with that, "today's faculty choose students on the basis of an array of perceptions that only sometimes have a strong evidentiary basis" (Posselt, 2016, p. 176). Therefore, professionalization of admissions staff and formation of communities of good admissions practices are required. Even despite certain gaps in research, already existing evidence allows significant progress toward the evidence-based policy on selective admissions for graduate schools across the world. Moreover, in addition to professionalization of admissions staff, it is important to consider monitoring and evaluation of the admissions process: Is there a closed-loop control of the admissions process? Are the selection methods scrutinized adequately in accreditation? Is there sufficient reporting on the chosen admissions process and selection methods applied in the HEI to higher levels? (e.g., Utrecht University, 2018). Ultimately, the answers to these questions reflect the extent of accountability of admissions committees for the soundness of their admissions practices. Accountability would imply reporting on data on each selection round to higher levels within HEI's organization. Institutional research, in turn, could have a role in analyzing emerging patterns, testing these against relevant models, and giving warning signals when substantial deviations occur. This would contribute to an adaptive admissions process that could eventually lead to fairer and more objective graduate admissions (Zimmermann et al., 2017).

2.4.5.3 Social Implications

Considering increasing numbers of applications and capacity limitations at research universities, evidence-based student selection is increasingly recognized as a socially significant practice which should diminish rather than enhance inequality. Failing to meet requirements of fairness, objectiveness, and transparency primarily leads to missed opportunities for capable students and a HEI, the inability of a HEI to justify the selection decisions, jeopardizing the diversity of the student body, infringement of students' rights on equal access to higher education, and the loss of time and efforts both by students and institutions. In extreme cases, abandoning quality requirements toward selective admissions process might lead to appearances of criminal bribing schemes (e.g., the 2019 college admissions bribery scandal in the USA). Designing a sound admissions process for graduate level education is, therefore, a necessary step for preventing these issues from arising or to cease their existence entirely. Finally, student selection has become an increasingly politicized societal topic, where advocacy groups and politicians are actively participating. In some countries, the alternatives to selective admissions are discussed, such as re-introducing the (weighted) lottery system in the Netherlands as a more neutral solution (The national government of the Netherlands, 2021). However, there is some critique of its effect on equal access, because a weighted lottery is based on selection criteria as well (Council of State of the Netherlands, 2021).

2.5 Conclusion

The main aim of this review was to collect, map, synthesize, and critically analyze the available research evidence on graduate selection methods with a focus on STEM disciplines. The results of the systematic search of research literature were categorized according to a type of selection method and core evaluative quality principles (predictive validity, acceptability, and procedural issues). Ten categories of graduate selection methods emerged. It was found that the predictive validity of prior grades, GRE General, intelligence assessments, and conscientiousness toward several study success dimensions is of medium-to-strong extent. Letters of recommendation, tests on language proficiency, emotional intelligence, and need for cognition are valid as well, but of weak-to-medium extent. Based on the limited evidence, it also appears that prior research experience, multiple mini-interviews, and selectivity of prior institution might have significant relationships with certain dimensions of graduate study success. Personal statements, traditional interviews, and personal traits such as extroversion and neuroticism are invalid predictors of graduate study success.

When choosing the selection methods to be applied in the admissions process, policy makers and admissions committees should use only valid instruments. They should also be aware of typical applicant reactions toward these methods as well as procedural issues such as possible adverse effects toward certain groups, susceptibility

for biases, faking, coaching, and stereotype threat. The admissions committees are advised (1) to completely exclude invalid selection instruments from their admissions requirements, (2) to define the dimensions of study success that are most important for their program, (3) to use those selection methods that showed predictive validity toward these predefined study success dimensions, accounting for applicant reactions and procedural issues of each of those methods, and (4) to ensure the accountability of the admissions process by reporting on data on each selection round to higher levels within HEI's organization, which should in turn conduct further analysis and regular evaluations of admissions processes.

2.6 References

Studies that met inclusion criteria and were included in the "Results" section of this review are marked with an asterisk.

Abunawas, M. E. (2016). A meta-analytic investigation of the predictive validity of the test of English as a foreign language (TOEFL) scores on GPA*. [Doctoral Dissertation, Texas A&M University]. The OAKTrust Digital Repository. https://hdl.handle.net/1969.1/154156

*Álvarez-Montero, F., Mojardin-Heraldez, A., & Audelo-Lopez, C. (2014). Criteria and instruments for doctoral program admissions. *Electronic Journal of Research in Educational Psychology*, *12*(3), 853–866. https://10.14204/ejrep.34.13138

Anderson, N., Salgado, J. F., & Hülsheger, U. R. (2010). Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment*, *18*(3), 291–304. http://dx.doi.org/10.1111/j.1468-2389.2010.00512.x

*Arcuino, C. L. T. (2013). The relationship between the Test of English as a Foreign Language (TOEFL), the International English Language Testing System (IELTS) scores and academic success of international master's students. [Doctoral dissertation, Colorado State University]. ProQuest LLC. https://www.proquest.com/docview/1413309058

*Attali, Y., & Sinharay, S. (2015). *Automated trait scores for GRE ® writing tasks: Automated trait scores for GRE ® writing tasks*. (Report No. RR –15- 15). Educational Testing Service. https://doi.org/10.1002/ets2.12062

*Biernat, M., & Eidelman, S. (2007). Translating subjective language in letters of recommendation: The case of the sexist professor. *European Journal of Social Psychology*, *37*(6), 1149–1175. https://doi.org/10.1002/ejsp.432

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures: Job applicant faking on personality measures. *International Journal of Selection and Assessment*, *14*(4), 317–335. https://doi.org/10.1111/j.1468-2389.2006.00354.x

*Bleske-Rechek, A., & Browne, K. (2014). Trends in GRE scores and graduate enrollments by gender and ethnicity. *Intelligence*, *46*, 25–34. https://doi.org/10.1016/j.intell.2014.05.005

*Boyette-Davis, J. (2018). A data-based assessment of research-doctorate programs in the United States. *The Journal of Undergraduate Neuroscience Education*, *17*(1), A54–A58. https://doi.org/10.17226/12994

*Bridgeman, B. (2016). Can a two-question test be reliable and valid for predicting academic outcomes? *Educational Measurement: Issues and Practice*, *35*(4), 21–24. https://doi.org/10.1111/emip.12130

*Bridgeman, B., Burton, N., & Cline, F. (2009). A note on presenting what predictive validity numbers mean. *Applied Measurement in Education*, *22*(2), 109–119. https://doi.org/10.1080/08957340902754577

*Briihl, D. S., & Wasieleski, D. T. (2007). The GRE analytical writing test: Description and utilization. *Teaching of Psychology*, *34*(3), 191–193. https://doi.org/10.1080/00986280701498632

*Burmeister, J., McSpadden, E., Rakowski, J., Nalichowski, A., Yudelev, M., & Snyder, M. (2014). Correlation of admissions statistics to graduate student success in medical physics. *Journal of Applied Clinical Medical Physics*, *15*(1), 375–385. https://doi.org/10.1120/jacmp.v15i1.4451

*Burton, N. W., & Wang, M. (2005). *Predicting long-term success in graduate school: A collaborative validity study.* (Report No. 99-14R. ETS RR-05-03). Educational Testing Service. http://grad.uga.edu/wp-content/uploads/2017/09/GRE_Research_Report.pdf

*Butter, R., & Born, M. Ph. (2012). Enhancing criterion-related validity through bottom-up contextualization of personality inventories: The construction of an ecological conscientiousness scale for Ph.D. candidates. *Human Performance*, *25*(4), 303–317. https://doi.org/10.1080/08959285.2012.703730

*Camara, W., Packman S., & Wiley A. (2013). College, graduate, and professional school admissions testing. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R., Kuncel, S. P., Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, Vol. 3. Testing and assessment in school psychology and education* (pp. 297–318). American Psychological Association. https://doi.org/10.1037/14049-014

*Chari, D., & Potvin, G. (2019). Admissions practices in terminal master's degree-granting physics departments: A comparative analysis. *Physical Review Physics Education Research*, *15*(1), Article 010104. https://doi.org/10.1103/PhysRevPhysEducRes.15.010104

*Cho, Y., & Bridgeman, B. (2012). Relationship of TOEFL iBT® scores to academic performance: Some evidence from American universities. *Language Testing, 29*(3), 421–442. https://doi.org/10.1177/0265532211430368

*Cline, F., & Powers, D. (2014). Test-taker perceptions of the role of the GRE® General test in graduate admissions: Preliminary findings. In *The Research Foundation for the GRE revised General Test: A Compendium of Studies* (p. 6.1.1-6.1.6). Educational Testing Service. https://www.ets.org/s/research/pdf/gre_compendium.pdf

Council of Graduate Schools. (2021). *CGS best practices programs in graduate admissions and enrollment management*. https://cgsnet.org/admissions-and-recruitment

Council of State of the Netherlands [Raad van State]. (2021). Amendment of the Higher Education and Scientific Research Act in relationship to the addition of decentralized draw as a selection method for higher education programs with fixed capacity [Wijziging van de wet op het hoger onderwijs en wetenschappelijk onderzoek in verband met het toevoegen van decentrale loting als selectiemethode voor opleidingen met capaciteitsfixus in het hoger onderwijs]. (W05.20.0508/I). https://www.raadvanstate.nl/@123920/w05-20-0508/

*Cox, G. W., Hughes Jr., W. E., Etzkorn, L. H., & Weisskopf, M. E. (2009). Predicting computer science Ph.D. completion: A case study. *IEEE Transactions on Education*, *52*(1), 137–143. https://doi.org/10.1109/TE.2008.921458

*Dabney, K. P. T. (2012). *Differences within: A comparative analysis of women in the physical sciences-motivation and background factors* [Doctoral dissertation, University of Virginia]. https://doi.org/10.18130/V3XZ9R

*Davey, T., & Lee, Y.-H. (2011). *Potential impact of context effects on the scoring and equating of the multistage GRE revised General test*. (Report No. GREB-08-01). Educational Testing Service. https://doi.org/10.1002/j.2333-8504.2011.tb02262.x

*DeClou, L. (2016). Who Stays and for How Long: Examining Attrition in Canadian Graduate Programs. *Canadian Journal of Higher Education*, *46*(4), 174–198. https://files.eric.ed.gov/fulltext/EJ1128619.pdf

*Dore, M. (2017). *Factors in the admissions process influencing persistence in a Master's of Science program in marine science* [Doctoral dissertation, Nova Southeastern University]. NSUWorks. https://nsuworks.nova.edu/cgi/viewcontent.cgi?article=1117&context=fse_etd

European Grade Conversion System. (2020). *Grade conversion–an introduction*. http://egracons.eu/page/about-egracons-project-and-tool

*Fischer, F. T., Schult, J., & Hell, B. (2013). Sex-specific differential prediction of college admissions tests: A meta-analysis. *Journal of Educational Psychology*, *105*(2), 478–488. https://doi.org/10.1037/a0031956

*Frasier, H. S. (2013). *An analysis of institutional characteristics that contribute to extended time to doctoral degree*. [Doctoral dissertation, University of Maryland]. *ProQuest LLC*. https://drum.lib.umd.edu/bitstream/handle/1903/14911/Frasier_umd_0117E_14734.pdf?sequence=1&isAllowed=y

*Fu, Y. (2012). *The effectiveness of traditional admissions criteria in predicting college and graduate success for american and international students*. [Doctoral dissertation, University of Arizona]. ProQuest LLC. https://www.proquest.com/docview/917742486

Garaz, S., & Torotcoi, S. (2017). Increasing access to higher education and the reproduction of social inequalities: The case of roma university students in Eastern and Southeastern Europe. *European Education*, *49*(1), 10–35. https://doi.org/10.1080/10564934.2017.1280334

*Garces, L. M. (2014). Aligning diversity, quality, and equity: The implications of legal and public policy developments for promoting racial diversity in graduate studies. *American Journal of Education*, *120*(4), 457–480. https://doi.org/10.1086/676909

*Gilmore, J., Vieyra, M., Timmerman, B., Feldon, D., & Maher, M. (2015). The relationship between undergraduate research participation and subsequent research performance of early career STEM graduate students. *The Journal of Higher Education*, *86*(6), 834–863. https://doi.org/10.1353/jhe.2015.0031

*Ginther, A., & Elder, C. (2014). A comparative investigation into understandings and uses of the TOEFL iBT® test, the international English language testing service (academic) test, and the Pearson test of English for graduate admissions in the United States and Australia: A case study: An investigation into test score understandings and uses. (Report No. RR– 14-44). Educational Testing Service. https://doi.org/10.1002/ets2.12037

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*(1), 26–34. https://doi.org/10.1037/0003-066X.48.1.26

*Hall, J. D., O'Connell, A. B., & Cook, J. G. (2017). Predictors of student productivity in biomedical graduate school applications. *PLOS ONE*, *12*(1), e0169121. https://doi.org/10.1371/journal.pone.0169121

*Hausknecht, J. P., Halpert, J. A., Paolo, N. T. D., & Gerrard, M. O. M. (2007). Retesting in selection: A meta-analysis of practice effects for tests of cognitive ability. *Journal of Applied Psychology*, *92*, 373–385. https://ecommons.cornell.edu/bitstream/handle/1813/76037/Hausknecht1_Retesting_in_Selction.pdf?sequence=1&isAllowed=y

*Howell, L. L., Sorenson, C. D., & Jones, M. R. (2014). Are undergraduate GPA and general GRE percentiles valid predictors of student performance in an engineering graduate program? *International Journal of Engineering Education*, *30*(5), 1145–1165. https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=2342&context=facpub

Jayakumar, U. M., & Page, S. E. (2021). Cultural capital and opportunities for exceptionalism: Bias in university admissions. *The Journal of Higher Education*, *92*(7), 1109–1139. https://doi.org/10.1080/00221546.2021.1912554

*Jones, J. M. (2013). *Evaluating intentionality: A psychometric rubric for statements of purpose*. [Doctoral dissertation, Grand Canyon University]. ProQuest LLC.

Kerrin, M., Mossop, L., Morley, E., Fleming, G., & Flaxman, C. (2018). Role analysis: The foundation for selection systems. In F. Patterson & L. Zibarras (Eds.), *Selection and recruitment in the healthcare professions: Research, theory and practice* (pp. 139–165). Palgrave Macmillan.

Kirby, W., & van der Wende, M. (2019). The New Silk Road: Implications for higher education in China and the West? *Cambridge Journal of Regions, Economy and Society*, *12*(1), 127–144. https://doi.org/10.1093/cjres/rsy034

*Klieger, D. M., Cline, F. A., Holtzman, S. L., Minsky, J. L., & Lorenz, F. (2014). *New perspectives on the validity of the GRE® General test for predicting graduate school grades: New perspectives for predicting graduate school grades*. (Report No. RR– 14-26). Educational Testing Service. https://doi.org/10.1002/ets2.12026

König, C. J., Steiner Thommen, L. A., Wittwer, A.-M., & Kleinmann, M. (2017). Are observer ratings of applicants' personality also faked? Yes, but less than self-reports. *International Journal of Selection and Assessment*, *25*(2), 183–192. https://doi.org/10.1111/ijsa.12171

*Kuncel, N. R., & Hezlett. S. A. (2007a). The utility of standardized tests: Response. *Science*, *316*, 1696–1697. https://doi.org/10.1126/science.316.5832.1694b

*Kuncel, N. R., & Hezlett, S. A. (2007b). Standardized tests predict graduate students' success. *Science*, *315*(5815), 1080–1081. https://doi.org/10.1126/science.1136618

*Kuncel, N. R., & Hezlett, S. A. (2010). Fact and fiction in cognitive ability testing for admissions and hiring decisions. *Current Directions in Psychological Science*, *19*(6), 339–345. https://doi.org/10.1177/0963721410389459

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, *86*(1), 148–161. https://doi.org/10.1037/0022-3514.86.1.148

*Kuncel, N. R., Kochevar, R. J., & Ones, D. S. (2014). A meta-analysis of letters of recommendation in college and graduate admissions: Reasons for hope. *International Journal of Selection and Assessment*, *22*(1), 101–107. https://doi.org/10.1111/ijsa.12060

*Kuncel, N. R., Wee, S., Serafin, L., & Hezlett, S. A. (2010). The validity of the Graduate Record Examination for master's and doctoral programs: A meta-analytic investigation. *Educational and Psychological Measurement*, *70*(2), 340–352. https://doi.org/10.1177/0013164409344508

Kuncel, N., Tran, K., & Zhang, S. H. (2020). Measuring student character: Modernizing predictors of academic success. In M. E. Oliveri & C. Wendler (Eds.), *Higher Education Admissions Practices: An International Perspective* (pp. 276–302). Cambridge University Press.

*Kurysheva, A., van Rijen, H. V. M., & Dilaver, G. (2019). How do admissions committees select? Do applicants know how they select? Selection criteria and transparency at a Dutch University. *Tertiary Education and Management*, *25*(4), 367–388. https://doi.org/10.1007/s11233-019-09050-z

Kyllonen, P. C., Walters, A. M., & Kaufman, J. C. (2011). *The role of noncognitive constructs and other background variables in graduate education.* (Report No. GREB-00-11). Educational Testing Service. https://doi.org/10.1002/j.2333-8504.2011.tb02248.x

*Kyllonen, P., Walters, A. M., & Kaufman, J. C. (2005). Noncognitive constructs and their assessment in graduate education: A review. *Educational Assessment*, *10*(3), 153–184. https://doi.org/10.1207/s15326977ea1003_2

*Lee, Y.-J., & Greene, J. (2007). The predictive validity of an ESL placement test: A mixed methods approach. *Journal of Mixed Methods Research*, *1*(4), 366–389. https://doi.org/10.1177/1558689807306148

Levashina, J., & Campion, M. A. (2007). Measuring faking in the employment interview: Development and validation of an interview faking behavior scale. *Journal of Applied Psychology*, *92*(6), 1638–1656. https://doi.org/10.1037/0021-9010.92.6.1638

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology*, *67*(1), 241–293. https://doi.org/10.1111/peps.12052

*Lorden, J. F., Ed, Kuh, C. V., Ed, & Voytuk, J. A., Ed. (2011). *Research-doctorate programs in the biomedical sciences: Selected findings from the NRC assessment*. Washington, DC: The National Academies Press. https://doi.org/10.17226/13213.

*Lott, J. L. I., Gardner, S., & Powers, D. A. (2009). Doctoral student attrition in the STEM fields: An exploratory event history analysis. *Journal of College Student Retention: Research, Theory and Practice*, *11*(2), 247–266. https://doi.org/10.2190/CS.11.2.e

*MasterMind Europe. (2017). *Admissions to English-taught programs (ETPs) at master's level in Europe–Procedures, regulations, success rates and challenges for diverse applicants*. ACA, StudyPortals, and Vrije Universiteit Amsterdam. http://mastermindeurope.eu/wp-content/uploads/2017/01/Report-2-Admissions-to-ETPs.pdf

*Mathews, J. (2007). Predicting international students' academic success... may not always be enough: Assessing Turkey's foreign study scholarship program. *Higher Education*, *53*(5), 645–673. https://doi.org/10.1007/s10734-005-2290-x

*Megginson, L. (2009). Noncognitive constructs in graduate admissions: An integrative review of available instruments. *Nurse Educator*, *34*(6), 254–261. https://doi.org/10.1097/NNE.0b013e3181bc7465

Merriam-Webster dictionary. (n.d.). Grade inflation. In *Merriam-Webster.com dictionary*. Retrieved February 18, 2022, from https://www.merriam-webster.com/dictionary/grade%20inflation

Miettinen, R. (2004). The roles of the researcher in developmentally-oriented research. In T. Kontinen (Ed.), *Development intervention. Actor and activity perspectives* (pp. 105–121). University of Helsinki, Center for Activity Theory and Developmental Work Research and Institute for Development Studies.

*Miller, A. (2013). *Timely doctoral completion rates in five fields: A two-part study*. [Doctoral dissertation, University of South Florida]. Digital Commons @USF. https://digitalcommons.usf.edu/cgi/viewcontent.cgi?article=6023&context=etd

*Miller, E. M. (2019). Promoting student success in statistics courses by tapping diverse cognitive abilities. *Teaching of Psychology*, *46*(2), 140–145. https://doi.org/10.1177/0098628319834198

*Moneta-Koehler, L., Brown, A. M., Petrie, K. A., Evans, B. J., & Chalkley, R. (2017). The Limitations of the GRE in predicting success in biomedical graduate school. *PLOS ONE*, *12*(1), Article e0166742. https://doi.org/10.1371/journal.pone.0166742

*Morgan, W. B., Elder, K. B., & King, E. B. (2013). The emergence and reduction of bias in letters of recommendation: Bias in letters of recommendation. *Journal of Applied Social Psychology*, *43*(11), 2297–2306. https://doi.org/10.1111/jasp.12179

*Mupinga, E. E., & Mupinga, D. M. (2005). Perceptions of international students toward GRE. *College Student Journal*, *39*(2), 402–409. https://go.gale.com/ps/i.do?id=GALE%7CA133606109&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=01463934&p=AONE&sw=w&userGroupName=anon%7E35cf9f02

*Murphy, K. R. (2009). How a broader definition of the criterion domain changes our thinking about adverse impact. In J. L. Outtz (Ed.), *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 137–160). Routledge. https://doi.org/10.4324/9780203848418

Murphy, S. C., Klieger, D. M., Borneman, M. J., & Kuncel, N. R. (2009). The predictive power of personal statements in admissions: A meta-analysis and cautionary tale. *College & University*, *84*(4), 83–88. https://www.proquest.com/docview/225605798

Okahana, H., & Zhou, E. (2018). *International graduate applications and enrollment: Fall 2017* (pp. 1–24). Washington, DC: Council of Graduate Schools. https://cgsnet.org/Data-Insights/

*oude Egbrink, M. G. A., & Schuwirth, L. W. T. (2016). Narrative information obtained during student selection predicts problematic study behavior. *Medical Teacher*, *38*(8), 844–849. https://doi.org/10.3109/0142159X.2015.1132410

*Pacheco, W. I., Noel, R. J., Porter, J. T., & Appleyard, C. B. (2015). Beyond the GRE: Using a composite score to predict the success of Puerto Rican students in a biomedical PhD program. *CBE—Life Sciences Education*, *14*(2), Article14:ar13. https://doi.org/10.1187/cbe.14-11-0216

*Park, H.-Y., Berkowitz, O., Symes, K., & Dasgupta, S. (2018). The art and science of selecting graduate students in the biomedical sciences: Performance in doctoral study of the foundational sciences. *PLOS ONE*, *13*(4), Article e0193901. https://doi.org/10.1371/journal.pone.0193901

Patterson, F., Roberts, C., Hanson, M. D., Hampe, W., Eva, K., Ponnamperuma, G., Magzoub, M., Tekian, A., & Cleland, J. (2018). 2018 Ottawa consensus statement: Selection and recruitment to the healthcare professions. *Medical Teacher*, *40*(11), 1–9. https://doi.org/10.1080/0142159X.2018.1498589

Patterson, F., & Ferguson, E. (2010). Selection for medical education and training. In T. Swanwick (Ed.), *Understanding medical education* (pp. 352–365). Wiley-Blackwell. https://doi.org/10.1002/9781444320282.ch24

Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., & Cleland, J. (2016). How effective are selection methods in medical education? A systematic review. *Medical Education*, *50*(1), 36–60. https://doi.org/10.1111/medu.12817

Pavlov, G., Maydeu-Olivares, A., & Fairchild, A. J. (2019). Effects of applicant faking on forced-choice and Likert scores. *Organizational Research Methods*, *22*(3), 710–739. https://doi.org/10.1177/1094428117753683

*Payne, D. (2015). A common yardstick for graduate education in Europe. *Journal of the European Higher Education Area*, *2*, 21–48. https://www.handbook-internationalisation.com/en/handbuch/gliederung/?articleID=566#/Beitragsdetailansicht/675/566

*Perez, K. (2011). *GRE as a Predictor of Graduate Student Success at a Hispanic Serving Institution of Higher Education* [Doctoral dissertation, Florida International University]. https://doi.org/10.25148/etd.FI11042602

*Petersen, S. L., Erenrich, E. S., Levine, D. L., Vigoreaux, J., & Gile, K. (2018). Multi-institutional study of GRE scores as predictors of STEM PhD degree completion: GRE gets a low mark. *PloS One*, *13*(10), Article e0206570. https://doi.org/10.1371/journal.pone.0206570

*Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*(2), 322–338. https://doi.org/10.1037/a0014996

*Posselt, J. R. (2014). Toward inclusive excellence in graduate education: Constructing merit and diversity in PhD admissions. *American Journal of Education*, *120*(4), 481–514. https://doi.org/10.1086/676910

Posselt, J. R. (2016). *Inside graduate admissions: Merit, diversity, and faculty gatekeeping*. Harvard University Press.

*Posselt, J. R. (2018). Trust Networks: A new perspective on pedigree and the ambiguities of admissions. *The Review of Higher Education*, *41*(4), 497–521. https://doi.org/10.1353/rhe.2018.0023

*Powers, D. E. (2017). Understanding the impact of special preparation for admissions tests. In R. E. Bennett & M. von Davier (Eds.), *Advancing Human Assessment*. *The Methodological, Psychological and Policy Contributions of ETS*. (pp. 553–564). Springer Open. https://doi.org/10.1007/978-3-319-58689-2

*Powers, D. E. (2005). *Effects of preexamination disclosure of essay prompts for the GRE analytical writing assessment*. (Report No. 01-07R). Educational Testing Service. https://doi.org/10.1002/j.2333-8504.2005.tb01978.x

Proudfoot, S., & Hoffer, T. B. (2016). Science and engineering labor force in the US. In L. Gokhberg, N. Shmatko, & L. Auriol (Eds.), *The science and technology labor force* (pp. 77–120). Springer International Publishing. https://doi.org/10.1007/978-3-319-27210-8

Reumer, C., & van der Wende, M. (2010). Excellence and diversity: The emergence of selective admissions policies in Dutch higher education. A case study on Amsterdam University College. *Research & Occasional Paper Series: CSHE.15.10*, 1–28. https://cshe.berkeley.edu/publications/excellence-and-diversity-emergence-selective-admission-policies-dutch-higher-0

*Rock, J. L., & Adler, R. M. (2014). A descriptive study of universities' use of *GRE®* General test scores in awarding fellowships to first-year doctoral students: A descriptive study of universities' use of *GRE®* scores. (Report No. RR– 14-27). Educational Testing Service. https://doi.org/10.1002/ets2.12027

Salmi, J., & Bassett, R. M. (2014). The equity imperative in tertiary education: Promoting fairness and efficiency. *International Review of Education*, *60*(3), 361–377. https://doi.org/10.1007/s11159-013-9391-z

*Sanford, R. (2009). *Predictors of international graduate student success in U.S. universities: analysis of noncognitive student variables and institutional characteristics*. [Doctoral dissertation, University of La Verne]. ProQuest LLC https://www.proquest.com/openview/0ca5804f25bd0ac8eeb599e839f8d656/1?pq-origsite=gscholar&cbl=18750&diss=y

*Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin*, *143*(6), 565–600. https://doi.org/10.1037/bul0000098

*Sealy, L., Saunders, C., Blume, J., & Chalkley, R. (2019). The GRE over the entire range of scores lacks predictive ability for PhD outcomes in the biomedical sciences. *PloS One*, *14*(3), Article e0201634. https://doi.org/10.1371/journal.pone.0201634

The Bologna Declaration. Joint declaration of the European Ministers of Education, June 19, 1999. http://www.ehea.info/page-ministerial-conference-bologna-1999

The national government of the Netherlands [Rijksoverheid]. (2021). *Loten voor studie zorgt voor kansengelijkheid* [*Lots for study ensures equality of opportunity*]. https://www.rijksoverheid.nl/actueel/nieuws/2021/03/12/loten-voor-studie-zorgt-voor-kansengelijkheid

*Trapmann, S., Hell, B., Hirn, J.-O. W., & Schuler, H. (2007). Meta-analysis of the relationship between the Big Five and academic success at university. *Zeitschrift Für Psychologie*, *215*(2), 132–151. https://doi.org/10.1027/0044-3409.215.2.132

Utrecht University. (2018). *Utrecht University education guidelines [Richtlijn onderwijs Universiteit Utrecht]*. https://www.uu.nl/en/education/education-at-uu/the-educational-model

van Ooijen-van der Linden, L., van der Smagt, M. J., Woertman, L., & te Pas, S. F. (2017). Signal detection theory as a tool for successful student selection. *Assessment & Evaluation in Higher Education*, *42*(8), 1193–1207. https://doi.org/10.1080/02602938.2016.1241860

*van Os, W. (2007). Selection to the master's phase at the binary divide, a Dutch case study. *Tertiary Education and Management*, *13*(2), 127–140. https://doi.org/10.1080/13583880701238365

Vrielink, J., Lemmens, P., & Parmentier, S. (2011). Academic freedom as a fundamental right. *Procedia–Social and Behavioral Sciences*, *13*, 117–141. https://doi.org/10.1016/j.sbspro.2011.03.009

*Walsh, M. J. (2020). Online doctoral student grade point average, conscientiousness, and grit: A moderation analysis. *Journal of Educators Online*, *17*(1). Advance online publication. https://www.thejeo.com/

Weedon, E. (2017). The construction of under-representation in UK and Swedish higher education: Implications for disabled students. *Education, Citizenship and Social Justice*, *12*(1), 75–88. https://doi.org/10.1177/1746197916683470

*Weiner, O. D. (2014). How should we be selecting our graduate students? *Molecular Biology of the Cell*, *25*(4), 429–430. https://doi.org/10.1091/mbc.e13-11-0646

*Weissman, M. B. (2020). Do GRE scores help predict getting a physics Ph.D.? A comment on a paper by Miller et al. *Science Advances*, *6*(23), Article eaax3787. https://doi.org/10.1126/sciadv.aax3787

*Westrick, P. A. (2017). Reliability estimates for undergraduate grade point average. *Educational Assessment*, *22*(4), 231–252. https://doi.org/10.1080/10627197.2017.1381554

*Willcockson, I. U., Johnson, C. W., Hersh, W., & Bernstam, E. V. (2009). Predictors of student success in graduate biomedical informatics training: Introductory course and program success. *Journal of the American Medical Informatics Association*, *16*(6), 837–846. https://doi.org/10.1197/jamia.M2895

*Wilson, M. A., DePass, A. L., & Bean, A. J. (2018). Institutional interventions that remove barriers to recruit and retain diverse biomedical PhD students. *CBE—Life Sciences Education*, *17*(2), Article 17:ar27. https://doi.org/10.1187/cbe.17-09-0210

*Wilson, M. A., Odem, M. A., Walters, T., DePass, A. L., & Bean, A. J. (2019). A model for holistic review in graduate admissions that decouples the GRE from race, ethnicity, and gender. *CBE - Life Sciences Education*, *18*(1), Article 18:ar7. https://doi.org/ 10.1187/cbe.18-06-0103

*Wollast, R., Boudrenghien, G., Van der Linden, N., Galand, B., Roland, N., Devos, C., De Clercq, M., Klein, O., Azzi, A., & Frenay, M. (2018). Who are the doctoral students who drop out? Factors associated with the rate of doctoral degree completion in universities. *International Journal of Higher Education*, *7*(4), 143–156. https://doi.org/10.5430/ijhe.v7n4p143

*Wongtrirat, R. (2010). *English language proficiency and academic achievement of international students: A meta-analysis*. [Doctoral dissertation, Old Dominion University]. ProQuest LLC. https://www.proquest.com/docview/744086570

*Zimmermann, J., Heinimann, H. R., Brodersen, K. H., & Buhmann, J. M. (2015). A model-based approach to predicting graduate-level performance using indicators of undergraduate-level performance. *Journal of Educational Data Mining*, *7*(3), 151–176. https://doi.org/10.5281/zenodo.3554733

*Zimmermann, J., von Davier, A. A., Buhmann, J. M., & Heinimann, H. R. (2017). Validity of GRE General test scores and TOEFL scores for graduate admissions to a technical university in Western Europe. *European Journal of Engineering Education*, *43*(1), 144–165. https://doi.org/10.1080/03043797.2017.1343277

Zimmermann, J., von Davier, A., & Heinimann, H. R. (2017). Adaptive admissions process for effective and fair graduate admissions. *International Journal of Educational Management*, *31*(4), 540–558. https://doi.org/10.1108/IJEM-06-2015-0080

Supplementary material

Table A1 | *Key Words Used in the Search in Literature Data Bases.*

Graduate, master*, doctoral, stud*, degree, school*, level, work, program*, admission*, thes*, education, select*, applica*, (previous or prior) *adj* (performance or achievement), determinant*, predictor variables, grade point average, UGPA, GPA, average grade, (standardized or aptitude or admission* or knowledge or language) *adj* test*, Graduate Record Examination*, GRE, Test of English as a Foreign Language, TOEFL, International English Language Testing System, IELTS, noncognitive *adj* (construct* or factor* or assessment or attribute*), interview*, MMI, recommendation letter*, SLOR, LOR, reference*, personality, personal* *adj* (factor* or trait* or assessment or measures), resume*, cv, curriculum vitae, interest inventor*, extracurricular activit*, biodata, (bio or personal or motivatio*) *adj* (questionnair* or letter or statement* or essa*), sjt, situational judgment test*, situational judgement test*, (selection or assessment) *adj* (centre or center), emotional intelligence, ei, work experience, self-efficacy, cognitive ability, intelligence, (student or study or school or degree or academic or graduate) *adj* (outcome* or record or attainment or achievement or success or accomplishment or performance or failure or progress* or satisfaction or persistence or completion), (student or study or school or degree or academic or graduate) *adj* (retention or attrition or dropout or withdrawal or withdrew or dismissal or "non completion" or "non-completion" or "time to degree"), valid*, reliab*, associat*, predict*, relat*, psychometric, bias*, faking o, adverse impact, acceptab*, feasib*, interview* perspective*, (stakeholder* or student or applicant) *adj* perception*, cost-effectiveness, cost-effective, cost effectiveness, cost, human resource.

Note. * refers to truncation, allowing finding the words with different endings; The *adj* operator finds two terms next to each other in the specified order.

Table A2 | *Number of Articles Relating to Each Selection Method and Evaluative Quality Principle under Consideration.*

Selection Methods	Evaluative Quality Principles				
	Validity/ Reliability	Procedural issues	Acceptability	Cost-effectiveness	Total
Prior grades	15	2	7	0	23
Standardized tests of academic abilities	29	14	11	1	50
Letters of recommendation	3	3	2	0	8
Interviews	2	0	1	0	3
Personal statements	2	0	3	0	5
Personality assessments	5	1	4	0	9
Intelligence assessments	2	1	1	0	3
Language proficiency	11	0	4	0	13
Prior research experience	5	0	3	0	8
Various (resumes, former (type of) higher education institution, amount and quality of research experience or composite scores)	9	1	5	0	14
Total Articles for Each Research Question	55	13	15	1	80

Note. Some articles included more than one selection method or addressed more than one evaluative quality principle. This is the reason why numbers in the last columns and the last raw do not equal to the total sum of articles reviewed.

Table A3 | *Summary of the Relevant Findings for Each Selection Methods.*

Note. Study Type abbreviations in the table
Q: Quantitative
L: Longitudinal
R: Review
M-A: Meta-analysis

Prior Grades

Table A3.1| *Prior Grades. Evidence on Predictive Validity and Reliability.*

Authors & Year	Study type	OECD Country	Academic field	Educational level	Type of prior grades examined
Bridgeman et al. (2009)	Q, L	US	biology, chemistry, education, English, experimental psychology, and clinical psychology	master's and doctoral	UGPA
Burmeister et al. (2014)	Q, L	US	medical Physics	master's and doctoral	UGPA
Burton et al. (2005)	Q, L	US	chemistry, biology	master's and doctoral	UGPA
Cox et al. (2009)	Q, L	US	computer science	doctoral	UGPA
Dabney (2012)	Q, L	US	chemistry, physics	doctoral	Average grade in undergraduate chemistry and physics courses
DeClou (2016)	Q, L	Canada	across different disciplines	graduate, not specified whether master's or doctoral or both	the first-year average
Dore (2017)	Q, L	US	marine sciences	master's	UGPA UMGPA
Fu (2012)	Q, L	US	across different disciplines	master's and doctoral	UGPA
Howell et al. (2014)	Q, L	US	mechanical engineering	master's	UGPA

	Dimensions of Graduate Study success examined	Summary of relevant findings/conclusions
	GGPA	UGPA is a valid predictor of GGPA
	-faculty ratings -course average -1st year GGPA, -performance on exit exam	Even though UGPA was not statistically correlated with study success dimensions, the authors state that this result might have been influenced by the narrow ranges of UGPA in already admitted group. Their conclusion is that after the applicants have been selected by UGPA (and GRE), further selection should be consider the other metrics shown to be correlated with study success.
	-GGPA -mastery of discipline -professional productivity -communication skills	UGPA has a stronger predictive validity in chemistry departments than in biology departments
	degree completion/dropout	A relationship between undergraduate GPA and study success of Ph.D. students has not been shown. The authors found, however, significant relationships between study success and a combination of three variable (the grade in the graduate algorithms course, performance in the graduate core courses as a whole (algorithms, computer architecture, software engineering, and operating systems), and whether or not the student wrote a thesis at the M.S. level).
	years to PhD completion	The average grade in undergraduate courses was not related to graduate time to doctoral degree
	degree completion/dropout	The first-year average is not a significant predictor of degree completion. The authors state that it might be that already enrolled students are academically strong and therefore, other factors play into their decision to dropout.
	degree completion/dropout	UMGPA is a significant predictor ($p<0.002$). UGPA is marginally predictive ($p<0.1$). The authors explain the latter finding by the fact that UGPA can be increased by an applicant taking courses with a higher probability of getting higher scores.
	first year GGPA	UGPA significantly predicted first-year GGPA
	-GGPA -time to graduation -advisor performance rating -publication rating	UGPA was found to be a significant positive predictor of GGPA and a significant negative predictor of duration of time to graduation. However, when the scores of GRE General are in the model, UGPA does not add to prediction of advisor performance rating or to publication rating

Authors & Year	Study type	OECD Country	Academic field	Educational level	Type of prior grades examined
Moneta-Koehler et al. (2017)	Q, L	US	biomedical	doctoral	UGPA
Park et al. (2018)	Q, L	US	biomedical	doctoral	UGPA
Willcockson et al. (2009)	Q, L	US	biomedical	master's and doctoral	UGPA
Wollast et al. (2018)	Q, L	Belgium	all disciplines	doctoral	UGPA
Zimmermann et al. (2015)	Q, L	Switzerland	computer sciences	master's	-UGPA -the third-year undergraduate GPA
Wistrick (2017)	M-A	US	across disciplines	after undergraduate	UGPA

Table A3.2 | *Prior Grades. Evidence on procedural issues.*

Authors & Year	Study type	OECD Country	Academic field	Educational level	Procedural issue
Posselt (2014)	Qualitative observational	US	STEM	doctoral	Grade inflation
Zimmermann (2018)	Q, L	Switzerland	Different technical specialties with the majority from engineering	master's	Imperfections of grade conversion

	Dimensions of Graduate Study success examined	Summary of relevant findings/conclusions
	-PhD completion/dropout -passing the qualifying exam -time to doctoral degree -number of presentations -number of first author peer-reviewed publications -obtaining a grant or fellowship -an introductory course of formal biomedical graduate training (first semester grades) -GGPA -faculty evaluations to manage the workload and write creatively	UGPA significantly predicted graduation with a Ph.D, first semester grades, GGPA, and faculty evaluations.
	-an introductory course of formal biomedical graduate training	There is a significant positive correlation between UGPA and performance on an introductory course
	-performance on an introductory course of formal biomedical graduate training	UGPA is a strong predictor of performance in the introductory course
	degree completion/dropout	UGPA predicts the higher rate of doctoral completion
	GGPA	UGPA, but especially the third-year UGPA are significant predictors of GGPA
	n/a	UGPA has relatively high reliability. The study used the longitudinal GPA data (across semesters) for 62,122 students from 26 institutions. The reliability estimates ranged between .60–.65 (semester GPAs), .75–.79 (annual GPAs), and .89–.92 (fourth-year cumulative GPAs).

	Summary of relevant findings/conclusions
	The author indicates that there is a lack of students from underrepresented backgrounds in highly selective programs already at undergraduate level. At the same time, grade inflation is a wide-spread phenomenon at undergraduate level. Setting the initial grades' threshold for graduate admissions high, excludes the students from underrepresented groups even more.
	In the study, UGPA provided only a small incremental validity over the GRE scores to predict GGPA. The authors explain this finding by their crude grade conversion scheme.

Table A3.3 | *Prior Grades. Evidence on acceptability.*

Authors & Year	Study type	OECD Country	Academic field	Educational level
Boyette-Davis (2011)	Q, cross-sectional	US	neuroscience	doctoral
Chari et al. (2019)	Q cross-sectional	US	physics	master's and doctoral
Kurysheva et al. (2019)	Q, cross-sectional	Netherlands	life and natural sciences	master's
Mastermind Europe (2017)	Q, cross-sectional	Europe	all disciplines	master's
Posselt (2014)	Qualitative observational	US	humanities, social sciences, natural sciences	doctoral

Graduate Record Examinations (GRE)

Table A3.4 | *Graduate Record Examinations (GRE) and other graduate standardized tests. Evidence on predictive validity.*

Authors & Year	Study type	OECD Country	Academic field	Educational level	Name of the test (section of the test)
Álvarez-Montero et al. (2014)	R	Mexico	across all disciplines, including STEM	master's	EXANI-III
Bridgeman (2016)	R	n/a	across all disciplines, including STEM	master's and doctoral	GRE General (A)
Bridgeman et al. (2009)	Q, L	US	biology, chemistry, education, English, and psychology	master's and doctoral	GRE General (V, Q)
Burmeister et al. (2014)	Q, L	US	medical physics	master's and doctoral	GRE General (V, Q, A) & GRE Subject (Physics)
Burton et al. (2005)	Q, L	US	biology, chemistry, education, English, and psychology	master's and doctoral	GRE General (V, Q)

	Summary of relevant findings/conclusions
	The average importance of prior GPA in admissions decisions was 5.49 (*SD* = 0.82) on a scale from 1 (not at all valued) to 7 (incredibly valued).
	In the prioritization of admission criteria, UGPA and math/ physics grades are the most important in admissions decisions
	UGPA and grades for relevant courses were found to be in top-10 of most important criteria. When specific criteria were combined, grades were found to be of a moderate importance in admission decisions, approximately on the same level and motivation factors and prior academic background and not significantly more important than research background of an applicant and cognitive ability.
	Prior grades are a wide-spread admission requirement to European masters programs. Both coordinators of master's programs and applicants the cut-off scores for GPA were indicated in the requirements of 40-50% of programs.
	The author suggests that the reasons behind high acceptability of grades lie in the fact that admissions committees members use them as means of identifying elite intellectual community, to which they refer themselves.

	Dimensions of Graduate Study success examined	Summary of relevant findings/conclusions
	degree completion	EXANI-III has no predictive power, as other admission instruments outperform this test.
	various, depending on a study included in the review	The review shows that GRE-A has a modestly strong correlation to study success. GRE-A may also balance the results of multiple-choice test. Such a balance contributes to fairer admissions. Due to these reasons, the authors suggest assigning a greater weight to GRE-A scores in admission decisions.
	first-year graduate GPA	The authors used an approach which divides students within a department into quartiles based on GRE scores and the percent of students in the top and bottom quartiles earning a 3.8 or higher GPA in their first year of graduate study was noted. Even after controlling for undergraduate GPA quartiles (i.e., looking at GRE quartile differences within GPA quartiles), substantial differences related to GRE quartile were found.
	-faculty ratings -program Exit Exam -scores on the most difficult courses in the graduate program -first-year graduate GPA	The scores on GRE-V & GRE-Q were positively related to scores in the most difficult courses in the program as well as to the program exit exam. Out of all sections, GRE-A correlated the most with the overall faculty ratings. GRE Subject (physics) was correlated with faculty ratings.
	-GGPA -faculty ratings	The authors find that the combination of GRE scores and UGPA strongly predicts cumulative graduate GPA and faculty ratings. They indicate that these results hold in each discipline.

Authors & Year	Study type	OECD Country	Academic field	Educational level	Name of the test (section of the test)
Cox et al. (2009)	Q, L	US	computer sciences	doctoral	GRE General (V, Q, A)
Dore et al. (2017)	Q, L	US	marine science	master's	GRE General (V, Q, A)
Frasier (2013)	Q, L	US	all disciplines, including STEM	doctoral	GRE General – average program scores
Fu (2012)	Q, L	US	biology, chemistry, education, English, psychology	master's and doctoral	GRE General (V, Q, A) & GRE Subject (chemistry & biology)
Hall et al. (2017)	Q, L	US	biomedical	master's and doctoral	GRE General (V, Q, A)
Howell et al. (2014)	Q, L	US	mechanical engineering	master's	GRE General (V, Q, A)
Klieger et al. (2014)	M-A	not reported (most likely, across countries)	education, engineering, English, biological and biomedical sciences, mathematics and statistics, psychology, health professions and clinical sciences, business and management.	master's and doctoral	GRE General (V, Q, A)
Kuncel et al. (2007 a)	R	not reported (most likely, across countries)	humanities, social sciences, biological sciences, physical sciences, mathematics, and professional graduate programs in management, law, pharmacy, and medicine	master's and doctoral	GRE General, GRE Subject and other, specific profession related standardized graduate tests
Kuncel et al., (2007 b)	Response	n/a	the same as above	the same as above	the same as above

	Dimensions of Graduate Study success examined	Summary of relevant findings/conclusions
	degree completion /attrition	The authors did not find the GRE scores to be predictive. They refer to it as an artefact and explain it be a restricted range of the GRE scores: the scores of already admitted students are already high and therefore they have limited power in distinguishing potentially successful students from non-successful.
	-graduation -time to degree	The GRE scores were not predictive for examined dimensions of study success
	time to degree	The author found a relatively weak negative relationship between average GRE scores and doctoral time to degree. The author infers that with the increase of selectivity of a program and levels of academic skills of admitted students, the degrees tend to be completed faster.
	first-year graduate GPA	GRE-V and GRE-Q weakly correlated with first-year GPA. GRE-A was not related to GGPA. GRE Subject was the strongest predictor amongst all GRE scales. The author draws attention to the latter finding, suggesting that knowledge in the content area best predicts graduate study success.
	-time to degree -graduate student productivity as measured by a number of student publications	The GRE scores were not correlated to number of publications. The authors state that the range of GRE scores was not very restricted in their sample and strongly advise against using GRE scores for admissions decisions to biomedical graduate programs.
	-time to degree -faculty ratings -GGPA -publication ratings defined as "a measure of a student's productivity in producing peer-reviewed publications of their work by the time of the thesis defense"	Among GRE sections, GRE-Q is a meaningful predictor. GRE-V was not found to be significant in relation to any dimensions of study success. GRE-A is correlated positively with GGPA(positively), faculty ratings (positively), and time to degree (negatively).
	GGPA	The three GRE sections are predictive of study success.
	-faculty ratings -citation count -research productivity -degree completion -qualifying exam -GGPA -1st GGPA	The authors state that in many primary studies on predictive validity of standardized tests, the reported correlations are weakened by statistical artifacts and attenuating factors such as the range restriction and unreliability in the study success measures. When meta-analytic methods are used to correct for these artifacts, standardized test scores are positively related to dimensions of study success. The authors highlight that even modest correlations are meaningful predictors.
	the same as above	The authors sequentially answer the critique to the article above, that was communicated in letters to the editor by three researchers.

Authors & Year	Study type	OECD Country	Academic field	Educational level	Name of the test (section of the test)
Kuncel & Hezlett (2010 a)	R	not reported (most likely, across countries)	all disciplines, including STEM	undergraduate and graduate	GRE General, GRE Subject and other, specific profession related standardized graduate tests
Kuncel et al., (2010 b)	M-A	not reported (most likely, across countries)	humanities, social science, life science, and math/physical science	master's and doctoral	GRE General (V, Q)
Lorden et al., (2011)	Q, L	US	biomedical	doctoral	GRE General (a weighted average of GRE-V and GRE-Q)
Lott et al. (2009)	Q, L discrete-time event history analysis	US	STEM	doctoral	GRE General (a combined score of Q & V) as an absolute and a relative measure. A relative GRE value greater than 1.0 indicates a higher than average individual GRE score compared to one's peers in the program; a relative GRE value lower than 1.0 indicates a lower than average GRE score relative to a program's peers.
Miller (2013)	Q, L	US	chemical engineering, physics, economics, English and neuroscience	doctoral	GRE General (a weighted average)
Moneta-Koehler et al. (2017)	Q, L	US	biomedical	master's and doctoral	GRE General (V, Q, A)
Park et al. (2018)	Q, L	US	biomedical	doctoral	GRE General (V, Q)

	Dimensions of Graduate Study success examined	Summary of relevant findings/conclusions
	-qualifying exam -research productivity -publication citations -faculty ratings -1st year GPA -cumulative GPA	The authors state that based on thousands of studies on hundreds of thousands of participants, the overarching inference is that test scores are positive predictors of various dimensions of study success. Test scores are not correlated strongly with motivationally determined outcomes (e.g., degree completion). Another conclusion is that tests, that are specific to particular disciplines (e.g., GRE Subject) are better predictors of study success than scores on tests, addressing broader academic skills.
	-1st year GPA -GGPA -faculty ratings	Across nearly 100 studies and 10,000 students, this study found that GRE scores predict the three dimensions of study success well for both masters' and doctoral students, with differences that ranged from small to zero. When the authors averaged the validity indicators of GRE-V and GRE-A toward dimensions of study success, the difference in validity estimates of the GRE toward study success was only .03: on master's level the estimate for GRE validity was.30 and on doctoral level— .27. The authors conclude that the GRE is a useful decision-making tool for programs on both master's and doctoral level.
	-time to degree -completion rates	The correlations between completion rates and average GRE scores were found to be low, and in several fields even negative, except for physiology.
	doctoral attrition	The first finding: doctoral students with higher total GRE scores had higher rates of attrition at their higher education institution. The second finding: odds of attrition are lower for the students with higher relative GRE scores than their peers in the same program. In other words, students, who are better than average in a program have lower attrition rates than those below the program average. According to the authors, relative-to-program measure may better capture the program-specific propensity for success.
	timely doctoral completion rate	GRE average score had a positive relationship with timely doctoral completion rates for physics and English, but not for chemical engineering, neuroscience, economics.
	-first semester grades -GGPA -faculty ratings -degree completion -time to degree -graduate student productivity -conference presentations -obtaining an individual grant or fellowship -passing the qualifying exam	The results of this study show that GRE scores did not predict several examined outcomes such as degree completion, passing the qualifying exam, time to degree, number of conference presentations, first author papers, or individual grants or fellowships. GRE scores predicted moderately first semester grades and weakly—GGPA and faculty evaluations.
	performance on a course "Foundations in Biomedical Sciences" as defined (Fail/ Pass/High Performer)	GRE-Q & GRE-V were mildly contributory in univariate analysis but did not remain significant in the adjusted model.

Authors & Year	Study type	OECD Country	Academic field	Educational level	Name of the test (section of the test)
Perez (2011)	A mixed-method study (sequential explanatory mixed design) In quantitative part: L)	US	life sciences, physical sciences, education, H humanities, social sciences	master's and doctoral	GRE General (V, Q)
Petersen et al. (2018)	Q, L	US	STEM (biological sciences, physical sciences, chemical sciences, computer and information sciences, engineering, geosciences, mathematical Sciences and related technology areas)	doctoral	GRE General (V, Q)
Sanford (2009)	Q, L (descriptive co-relational ex post facto study)	US	engineering, computer science, physical sciences, social sciences, humanities, business, medicine, law, education, other	master's and doctoral (only international)	GRE General (V, Q, A)
Sealy et al. (2019)	Q, L	US	biomedical	master's and doctoral	GRE General (V, Q)
Miller et al. (2019)	Q, L	US	physics	doctoral	GRE General (V, Q) GRE Subject (Physics)
Weissman (2020)	Response on Miller et al. 2019	US	physics	doctoral	GRE General (V, Q) GRE Subject (Physics)
Willcockson et al. (2009)	Q, L	US	biomedical	master's and doctoral	GRE General (V, Q)
Zimmermann et al. (2018)	Q, L	Switzerland	STEM (the majority) & management, humanities, and social sciences (the minority)	master's	GRE General (V, Q, A)

	Dimensions of Graduate Study success examined	Summary of relevant findings/conclusions
	-GGPA -graduation rate (defined as completion of the degree with 6 years for master's students and 9 years for doctoral students)	It was concluded that GRE-Q and GRE-V predicted both dimensions of study success for master's students. GRE-Q did not predict graduation rate for doctoral students, though it predicted their GGPA.
	degree completion/attrition time to degree	GRE-V and GRE-Q scores did not differ for females who completed and those who left programs. Males in the lower quartiles of GRE-V or GRE-Q completed their degrees more often than males in the highest quartile. This pattern was stable across the four institutions in the study. GRE scores did not predict graduate time to degree.
	-GGPA (self-reported) -number of authored publications (self-reported)	The estimates for the GRE General did not reach the level of significance. The author note that the sample lacked variability in GRE.
	-time to degree -faculty evaluations -publications -first author publications -predoctoral fellowship -awards	The authors did not find GRE scores to be predictive to study success and the long-term graduate outcomes. This study used the data from a university, which historically admits underrepresented students with a wide span of GRE scores. The authors indicate that this allowed them to avoid the typical biases of other studies with narrow ranges in GRE scores and therefore their findings can be generalized to the populations of applicants.
	-degree completion	The study concluded that the GRE scales are not predictive for completion of PhD in physics. Completion was significantly related to GRE-Q in two of four studied models. GRE-V and GRE Physics were not predictive of completion in any model.
	-degree completion	Weissman identifies pitfalls of the study of Miller et al. (2019). H indicates several issues, among which: variance inflation by collinearity and range restriction, not completeness of a correlation matrix, inflation of the confidence intervals in a presented figure and others.
	performance in an introduction informatics course (Mastery or Failure grade)	GRE-V, but not GRE-Q was found a study success measure. GRE-Q was strong predictor of overall program performance.
	-GGPA -rate of progress (defined as number of credits obtained in the master's program divided by the number of study semesters until completion) -master's thesis performance	The authors show that the GRE-Q was a weak predictor of GGPA and study progress (and it did not predict thesis performance). Nevertheless, the authors state that their evidence indicates that the GRE General is a sensible admission instrument.

Table A3.5 | *Graduate Record Examinations (GRE) and other graduate standardized tests. Evidence on procedural issues.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level	Name of the test (section of the test)
Bleske-Rechek et al. (2014)	Q, L	US	across all	master's and doctoral	GRE General (V, Q)
Camara et al. (2010)	R	n/a	n/a	n/a	Different standardized admission tests, both undergraduate and graduate, including the GRE.
Davey et al., (2011)	Q, cross-sectional	US	across all disciplines	master's and doctoral	GRE General (V, Q)
Fischer et al., (2013)	Meta-Analysis	n/a	across all disciplines	master's and doctoral	Different standardized admission tests, both undergraduate and graduate, including the GRE, and subject-specific admission tests in German-speaking countries.
Garces (2014)	Review	n/a	n/a	n/a	Different standardized admission tests, including GRE

Procedural issue	Summary of relevant findings/conclusions
Gate-keeping role in terms of gender and ethnicity	The authors found that 1) The gap between males' and females GRE-Q scores has not substantially diminished since 1980-s . Along with that, female representation in STEM graduate programs has increased substantially. 2) Second, there is persistence of ethnic gaps on GRE performance. However, again, representation of historically disadvantaged ethnic groups in graduate programs has increased. The authors conclude that the efforts toward more equal access to graduate education were successful, even though GRE is still used in many HEIs.
1-performance of underrepresented minority students and female students, 2-role of SES, 3-family background and accumulated experience as a privilege in standardized tests performance, 4-impact of coaching	1– The author brings attention to the fact that the differences in performance on standardized tests have been pointed out as evidence of bias against underrepresented groups (minorities or females). The author states, however, that this difference in performance represents a reflection of the difference in educational opportunities, which different ethnic/gender groups have. The prior reviews and individual studies, according to the author, found that admissions test scores slightly over predict the performance of underrepresented minority students. It appears that standardized tests slightly underpredict females' study success. The author hypotheses that female students are more likely to study humanities or social sciences. Students, studying these disciplines, tend to have higher GGPAs than their peers who study natural sciences. Another hypothesis which the author gives is that females can organize themselves better, which contributes to their higher grades during studies than what have been predicted based on their performance on a test. 2– another criticism for the standardized admission tests is that the testing adds no value beyond a what SES predicts. The author reviews the meta-analyses that provided evidence that standardized (undergraduate) admission tests have predictive validity beyond SES. 3– the author reviews literature that found positive small to medium correlations of standardized test scores and parental education/family income. 4– the author states that there is no evidence to support the positive impact of coaching on test performance.
Item position effects of the revised GRE	The authors conclude that in order to mitigate position effects in multistage testing, the organizers of the testing should use put a special attention to the time limits and GRE questions configuration.
Test fairness and test bias in predicting subgroups	The meta-analysis found small underprediction of women's academic performance by standardized tests and the overprediction of men's academic performance. This means that females tend to earn undergraduate grades that are higher than those predicted, and in contrast, males tend to earn grades that are lower than predicted. The advice of the researchers is to combine admission tests with indicators of previous academic achievements, such as prior grades, because it = reduces the amount of under- and overprediction.
A negative effect of stereotype threat on standardized test performance	Students, who believe that their nonoptimal performance on standardized tests might confirm the stereotypes of their racial group's intellectual capacity, might perform worse on the standardized tests because of this perception.

Authors & Year	Study type	OECD Country	Academic field	(Educational) level	Name of the test (section of the test)
Hausknecht et al. (2007)	Meta-analysis	Not specified	not specified	not specified	Different standardized admission tests, including GRE
Kuncel et al. (2007 a)	R	not reported (most likely, across countries)	humanities, social sciences, biological sciences, physical sciences, mathematics, and professional graduate programs in management, law, pharmacy, and medicine	master's and doctoral	GRE General, GRE Subject and other, specific profession related standardized graduate tests
Kuncel & Hezlett (2010a)	R	not reported (most likely, across countries)	across all	all levels, undergraduate, when talking about the role of SES.	Different standardized admission tests, including GRE
Murphy (2009)	Qualitative	n/a	n/a	n/a	Different standardized admission tests, including GRE
Posselt (2014)	Qualitative	US	STEM, social sciences, humanities	doctoral	GRE General. It is not specified whether GRE Subject was also included
Powers (2017)	R	US	not specified	master's and doctoral	GRE General (V, Q, A)
Wilson et al. (2018)	Q, cross-sequential	US	STEM	doctoral	GRE General (V, Q)
Wilson et al. (2019)	Q,L	US	STEM	doctoral	GRE General (V, Q)

Procedural issue	Summary of relevant findings/conclusions
Retesting in selection/ coaching effects	Retesting has an effect on performance on standardized tests: the performance improves on approximately one quarter of a standard deviation when assessed from the first to the second time taking the test and on one-fifth of a standard deviation from the second to the third time of taking the test. The authors indicate that additional research is needed to understand the impact of retesting on the validity of test scores.
-Bias in testing -Coaching effects	Regarding bias in training, the authors state that regression lines in primary studies frequently do not differ by race or ethnic group. In those case when they differ, standardized tests favor underrepresented groups. The authors also find that tests tend to underpredict female undergraduate study success, but not graduate one. Regarding coaching effects in testing, the authors state that the effects are modest: around 25% of standard deviation. The further increments of improvement require substantially more efforts. The test developers put a special attention to eliminating those items, which were found to be susceptible to coaching. Finally, the authors state that test preparation or retaking does not change predictive validity of admissions tests.
-Role of SES -Predictive bias	Role of SES: The authors re-interpret the data from existing studies and show that (undergraduate) test scores are not just a proxy for SES, but they stay predictive even after SES and prior grades are controlled for. The authors argue that the differential performance on standardized tests reflects the societal problem, since certain groups receive less family, societal, environmental, peer and financial support to develop the (graduate) skills. The standardized tests only reflect these initial differences. Stopping the usage of tests will not resolve the origins of this societal problem.
Different performance of different demographic groups on standardized tests, more specifically adverse impact against members of lower-scoring groups	The author shows that it is difficult to accomplish two goals (of predicting performance and avoiding adverse impact) at the same time. The author proposes the steps that simplify decision-making in selective admissions. Following these steps, the decision makers are "forced" to think thoroughly of the adverse impact of the measures they use and to decide what they prioritize.
Misuse of GRE scores	The author finds that high GRE scores are included into the conceptualization of merit and are used to quickly compare students. The high standards in GRE scores represented for admissions committees the conventional achievement that are typical to obtain in an intellectual community (to which they refer themselves as well).
Test preparation	For GRE-A, there was a strong relationship revealed between the effect of coaching and its duration. However, coaching for GRE-Q and GRE-V did not show such a relationship, or extremely small one. These findings are in contrast to commercial companies which state that such relationship is strong
Adverse impact of GRE scores on program diversity	After the holistic and rolling approach in admissions started being implemented and the admission committees shifted the focus from quantitative metrics, the diversity of students who are offered admission, accept the offer and matriculate into the grad school increased without substantially affecting U-GPA or GRE scores of entering students.
Misuse of the GRE scores in metrics-based admission approach	Two approaches: metrics-based and holistic were compared. Metrics-based review of applicants excluded twice the number of applicants who identified as underrepresented groups compared with their peers. Using holistic approach, delivered more diverse student population. The authors state that applicant assessments in their holistic review process were independent of gender, racial, and citizenship status.

Table A3.6 | *Graduate Record Examinations (GRE) and other graduate standardized tests. Evidence on acceptability.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level	Name of the test (section of the test)
Attali et al. (2015)	Q, cross-sectional	US	n/a	n/a	GRE-General (A)
Boyette-Davis (2011)	Q cross sectional	US	neuroscience	doctoral	GRE (Type not specified)
Briihl et al. (2007)	Qualitative	US	psychology	master's and doctoral	GRE General (A)
Chari et al. (2019)	Q cross-sectional	US	physics	master's and doctoral	GRE General (V, Q, A) & GRE Subject
Cline et al. (2014)	Q cross-sectional	US	all disciplines	master's and doctoral	GRE General
Mastermind Europe (2017)	Q cross-sectional	Europe	all disciplines	master's	Admission tests in general
Mupinga et al. (2005)	Qualitative	US	different disciplines, not specified	international graduate students, level not specified	GRE General
Payne (2015)	Review	Europe	all disciplines	not defined, assumingly both master's and doctoral	GRE General
Powers (2005)	Qualitative	not reported	not reported	master's and doctoral	GRE-A
Rock et al. (2014)	Qualitative	US	STEM, humanities, social sciences	doctoral	GRE- General

	Summary of relevant findings/conclusions
	The study identified four trait scores: word choice, grammatical conventions, fluency, and organization. It was shown that applicants are usually very interested in receiving additional information on their performance beyond the total test score. Giving the scores on these four traits of their writing might increase the acceptability of GRE-A.
	The results indicate that the GRE scores are an important admissions criterion, but far not the most important one. Having previous research experience, U-GPA, and some personal characteristics (e.g., critical thinking, work ethics etc.) are more important for the admission committees.
	Only 35% of responded graduate programs use GRA-A in their admission decisions. Most of other programs (which do not use GRE-A), do not plan to start using it. Those programs, which use GRE-A rated it as medium or low in importance in their admission decisions. Only few programs use cut-off scores.
	GRE-Q and GRE-Subject are important criteria in admission decisions to physics programs, but they are less important for master's programs in comparison with doctoral programs. The authors indicate that applicants to masters' programs without highest scores on GRE-Q, may still have chances to be admitted. Along with that, GRE-V is a low rated factor in decision-making for the examined programs.
	White test takers have fewer negative feelings about the GRE General than racial minority test takers. There were only small differences found in the levels of anxiety and preparedness for the GRE test between these groups. Overall, most test takers view GRE as less important in graduate admission than UGPA, recommendation letters, and work experience.
	The survey of Mastermind Europe across masters' programs coordinators found out that that admission tests are used rarely as an eliminative requirement (only 5.3% programs mentioned this). When present, however, the results of admission tests provide an advantage in the selection process in 23.5% of master's programs that responded to the survey.
	International graduate students perceive that the GRE General has the issues with the content (such as that it tests unfamiliar concepts in GRE-V), the context (not being familiar with some practices that are not common outside of the US, e.g., baseball game), the purpose (according to the study participants, the GRE does not tape cognitive aspects, but a piece of knowledge that one learns from their culture and environment). Overall, the interviewed international students felt that the GRE is culturally biased and does not measure cognitive capacity. They would prefer that the GRE scores would be used together with other information about them such as UGPA, special talents, their accomplishments, recommendations, and interviews.
	The author indicates that due to reforms of European higher education and increased internationalization of graduate programs, many higher education institutions are now pursuing to standardize their graduate admission process. Payne provides evidence on growing number of GRE tests, undertaken by European students, also those who intend studying engineering, physical sciences, and life sciences were the largest groups across several years.
	The perceptions of applicants on publishing GRE-A essay prompts were positive.
	The authors conducted structured interviews with deans/directors, faculty members and administrators to identify the role of the GRE General in awarding fellowships to first-year doctoral students. The authors concluded that the GRE scores are applied as a method to understand whether applicants have the baseline knowledge and skills for a graduate program.

Table A3.7 | *Graduate Record Examinations (GRE) and other graduate standardized tests. Evidence on cost-effectiveness.*

Authors & Year	OECD Country	Academic field	(Educational) level	Name of the test (section of the test)
Klieger et al. (2014)	not reported (most likely, across countries)	education, engineering, English, biological and biomedical sciences, mathematics and statistics, psychology, health professions and clinical sciences, business, and management.	master's and doctoral	GRE General (V, Q, A)

Letters of Recommendations (LoRs)

Table A3.7 | *References, or Letters of Recommendations (LoRs). Evidence on predictive validity and reliability.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level	Format of recommendation letters
Hall et al. (2017)	Q, L	US	biomedical	master's and doctoral	A letter itself & Ratings in a letter on a 5-point Likert scale from "Below Average" to "Exceptional".
Kuncel et al. (2014)	Meta-analysis	Not specified	across different fields	undergraduate, graduate, and professional school	Different formats
Megginson (2009)	Review	Not specified	across different fields	graduate	Different formats, including Five-Trait Category Tool (FTCT) to standardize the evaluation of Narrative Letter of Recommendation (NLoR) and standardized letters of recommendation (SLORs)

	Summary of relevant findings/conclusions
	The author calculates the financial gains from using the GRE test. He takes validity coefficients for the GRE-V (0.19), GRE-Q (0.24), and GRE-A (0.21), respectively and calculates that the annual gain (without consideration of costs) a doctoral graduate program would be $55,940 (from GRE-V), $70,661 (from GRE-0Q), and $61,828 (from GRE-A). He also provides the estimates of financial gains if unadjusted range variation would be used.

	Dimensions of study success examined	Summary of relevant findings/conclusions
	-time to degree -graduate student productivity as measured by a number of student publications	The scores on LoRs did not predict time to degree. However, they were the most powerful predictor of number of student research publications. Students with more than three published articles had higher average scores on their LoRs than the students who did not have any publications.
	-GGPA -degree attainment -faculty ratings	The correlations of LoRs with faculty ratings and GGPA were found to be negligible. The correlation of LoRs with degree attainment, turned out to be small, however they provided incremental validity. This validity was superior or equal to all other traditional predictors (incl. standardized tests, field specific knowledge such as GRE-subject tests. The authors note that these results may indicate that further improvement in standardization (very much like with interviews) could be feasible.
	varies per study reviewed, however the author notes that the primary studies were often not rigorous on study success dimensions they aimed to assess with the noncognitive tools.	NLOR display minimal reliability, are prone to subjective interpretation. The review provides an overview of other noncognitive assessments that have a potential to become psychometrically sound selection methods. In spite of this potential, they are not sufficient at the moment in determining the core attributes that predict graduate study success.

Table A3.8 | *References, or Letters of Recommendations (LoRs). Evidence on procedural issues.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level	Format
Biernat & Eidelman (2007)	Q	US	physics	letters for graduate level; participants in the studies – undergraduate students	narrative
Morgan et al. (2013)	Q	US	psychology	letters for graduate level; participants in the studies – undergraduate students	narrative
Posselt (2018)	Qualitative	US	not specified	graduate, participants – the selection committee members themselves.	narrative

Table A3.9 | *References, or Letters of Recommendations (LoRs). Evidence on acceptability.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level	Format
Boyette-Davis (2011)	Q	US	neuroscience	doctoral	any
MasterMind Europe (2017)	Q	Europe	all disciplines	master's	any

Interviews

Table A3.10 | *Interviews. Evidence on predictive validity.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level	Level of structure of interviews
Hall et al. (2017)	Q, L	US	biomedical	master's and doctoral	Unstructured with an overall rating of the interview afterwards
oude Egbrink et al. (2016)	Q, L	The Netherlands	physician-clinical investigator	four-year research master's	Notes based on MMI

Procedural issue	Summary of relevant findings/conclusions
Gender bias	The gender of applicants was found to play a role in how the same favorable language is interpreted by admissions committees. They write that admissions committees assume that the same favorable language indicates less positive when it is used to describe a female versus a male
Gender bias Race bias	The authors state that requiring referees to expand on their scores in LoRs (i.e., explain) reduces the bias that is integrated into the scoring system of LoR" Explanation of scores in this study was done in 2 aspects: 1.They evaluated the target on four areas of performance: competence, culture, liking, and organizational citizenship behaviors. 2. Explaining why the certain rating for the LoR was chosen by a rater. A worthy side-finding is that the name alone on LoRs negatively influences raters' evaluation of underrepresented minority groups.
-Interpretation of letters based on the extent of familiarity with the letter writer. -Interpretation of positive language in a reserved tone - Cultural difference in the degree of exaggeration	-The author notes that trusting relationships with letter writers aided professors in interpreting letters of recommendation -The author observed that when a letter was written in a positive language in a reserved tone, "committees debated whether the tone should be interpreted as indicative of the writer's personality or as a lack of enthusiasm about the applicant" -The author notes that American LoRs have become very inflated, and it is also recognized by the selection committees.

Summary of relevant findings/conclusions
LoRs were found in this study to be the second most valuable component. Admissions committee's comments indicated that could infer information about personal characteristics (such as critical thinking, reliability etc.)
The survey of Mastermind Europe across masters' programs coordinators found out that that LoR(s) were not explicitly mentioned by most of master's coordinators and fell apparently in the category "other criteria". They have not been mentioned in reasons for failure in the admissions process either. The applicants, however, indicated that they were asked to submit one or more LoRs in 58.3% cases.

Dimensions of study success examined	Summary of relevant findings/conclusions
graduate student productivity as measured by: -time to degree -several student publications	There was no relationship found between faculty interview ratings and high or low student productivity.
problematic study behavior (such as planning and self-reflection)	A significant correlation was found between the notes after interviews and planning problems. No such evidence was found for self-reflection related problems. The authors conclude that information in the narrative format noted during MMIs contains predictive indicators for planning-related issues of students. This could be used for identification of students at risk and their further counselling.

Table A3.11 | *Interviews. Evidence on acceptability.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level	Level of structure of interviews
MasterMind Europe (2017)	Q, survey	Europe	all disciplines	master's	not specified

Personal Statements, or Motivation Letters

Table A3.12 | *Personal Statements, or Motivation Letters. Evidence on predictive validity.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level	Structure of personal statements
Jones (2013)	Psychometric, mixed-methods study	US	not specified	doctoral	Unstructured. The 5-point intentionality rubric was used to evaluated them
Murphy et al. (2009)	Meta-analysis	not specified	not specified	master's and doctoral	Mostly unstructured

Table A3.13 | *Personal Statements, or Motivation Letters. Evidence on acceptability.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level	Format
Kurysheva et al. (2019)	Q	Netherlands	life and natural sciences	master's	unstructured
MasterMind Europe (2017)	Q	Europe	all disciplines	master's	any
Posselt (2014)	Qualitative	US	STEM, social sciences, humanities	doctoral	unstructured

	Summary of relevant findings/conclusions
	Interview is used in a 22.6% of English-taught masters' programs across Europe. There are indications that master's coordinators wish to use it more as they value additional information about the applicant they infer from interviews. Poor interview is a reason for rejection of an applicant in less than 5% cases.

	Dimensions of study success examined	Summary of relevant findings/conclusions
	none	A tool was developed to measure intentionality in personal statements. It measured four aspects: motivation, beliefs, goal setting, self-efficacy. The intentionality rubric, developed by the author, allows HEIs to give scores on intentionality. In this study, the first evidence for the construct validity of the rubric was shown. The rubric would need to be tested further to establish its predictive value for graduate study success.
	-first-year graduate GPA and GGPA combined -faculty ratings	While personal statements have little overlap with other predictors, they also have small predictive value toward grades and faculty performance. Most importantly, personal statements do have incremental validity.

	Summary of relevant findings/conclusions
	Motivation letters are amongst most frequently used tools for assessment of admission criteria. They are used to assess not only motivation itself, but also criteria related to personality, cognitive ability, and previous academic background.
	The survey of Mastermind Europe across masters' program coordinators found that motivation letters are reported to be required during admission in 51.5% of European master's programs and their better evaluation by selection committees provides advantage in 23.5% cases. Poor motivation letter almost never serves as a reason for failure (only in 5% of cases, according to master's coordinators). The applicants report on similar usage of motivation letters (in 59.1% cases). Another finding related to the fact that admissions committees require a motivation letter from foreign applicants twice as often as from internal students.
	Admissions committees admitted that they do not trust to the narratives in the letters, and they also do not have enough time to read all of them in detail. If there GRE scores are not good, most likely the letters won't be read. They do play a role, when a committee has a specific goal of diversity, and they try to find evidence that a student bring a unique and valuable perspective.

Intelligence Assessments

Table A3.14 | *Intelligence Assessments. Evidence on predictive validity.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level	The intelligence test examined
Poropat (2009)	M-A	across countries	across disciplines	tertiary education	across different intelligence assessments
Schneider et al. (2017)	R of M-A	across countries	across disciplines	tertiary education	across different intelligence assessments

Table A3.15 | Intelligence Assessments. Evidence on procedural issues.

Authors & Year	Study type	OECD Country	Academic field	(Educational) level	Procedural Issues
Poropat (2009)	M-A	across countries	across disciplines	tertiary education	overlap with other popular selection criteria

Table A3.16 | Intelligence Assessments. Evidence on acceptability.

Authors & Year	Study type	OECD Country	Academic field	(Educational) level
Kurysheva et al. (2019)	Q, cross-sectional	Netherlands	life and natural sciences	master's

Personality Assessments

Table A3.17 | *Personality Assessments. Evidence on predictive validity.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level	Type of personality assessment
Butter et al. (2012)	Q, L	The Netherlands	not specified	doctoral	-Ecological Conscientiousness or time management scale for Ph.D. candidates (examples of the items: "I can keep myself going", "my time management in my research is very accurate and realistic") -the Big Five -more specific instruments (frame-of-reference scales that are more situation-specific than the Big Five; narrow trait scales that are more trait-specific than the Big Five).

Dimensions of study success examined	Summary of relevant findings/conclusions
GPA	Intelligence has a correlation of .23 with academic performance and d=0.52. This correlation is a similar size as for Conscientiousness. The authors infer that it is likely to have similar levels of practical utility.
academic achievement, defined as procedural and declarative knowledge and as grades or performance on an educational achievement test	Intelligence is one of the main correlates of the academic achievement, along with high self-efficacy, high prior achievement, and conscientious. It has medium-to-large effect size.

Summary of relevant findings/conclusions
The practical usefulness of intelligence as a predictor of future study success is limited because it has a substantial overlap with previous academic performance. In real world, the admissions committees cannot correct for various artefacts including range-restricted groups of applicants.

Summary of relevant findings/conclusions
On two samples of students in the life sciences and the natural sciences, it was found that admission criteria, related to cognitive ability, are regarded to be important in admission decisions by the admission committees, though they do not make it to the top 10 of most important specific selection criteria. It was also found that information about cognitive ability is often inferred by the admission committees members not only from the cognitive tests, but also from other admission documents (e.g., a motivation letter, interviews, recommendation letter, personal acquaintance)

Dimensions of study success examined	Summary of relevant findings/conclusions
-research progress, -meeting deadlines, -probability to obtain the Ph.D. in time	The developed scale significantly predicted Ph.D. performance criteria and showed incremental validity beyond Big Five and more specific instruments like frame-of-reference scales and situation-specific

Authors & Year	Study type	OECD Country	Academic field	(Educational) level	Type of personality assessment
Poropat (2009)	M-A	across countries	across disciplines	tertiary education	Big Five traits
Schneider et al. (2017)	R of M-A	across countries	across disciplines	tertiary education	Different personality assessments, depending on a meta-analytical study under examination.
Trapmann et al. (2007)	M-A	across countries	across disciplines	tertiary education	Big Five
Walsh (2020)	Q, cross -sectional	US	across disciplines	doctoral (online program)	-Grit scores, self-reported (using Grit-S scale) -Conscientiousness scores, self-reported (using the Big Five)

Table A3.18 | *Personality Assessments. Evidence on procedural issues.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level	Procedural Issues
Kyllonen et al. (2005)	R	across countries	across disciplines	graduate	-Faking -coachability

Table A3.19 | *Personality Assessments. Evidence on acceptability.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level
Boyette-Davis (2011)	Q, cross-sectional	US	neuroscience	doctoral
Kurysheva et al. (2019)	Q, cross-sectional	Netherlands	life and natural sciences	master's
Kyllonen et al. (2005)	R	across countries	zcross disciplines	graduate

Dimensions of study success examined	Summary of relevant findings/conclusions
GPA	Study success had relationships with Agreeableness, Conscientiousness, and Openness. Importantly, correlations between Conscientiousness and study success were independent of intelligence. When study success in secondary school was controlled for, Conscientiousness and Intelligence had similar effect sizes.
academic achievement defined as procedural and declarative knowledge and as grades or performance on an educational achievement test	Students with high achievements are characterized by several characteristics, including conscientiousness. Absolute effect sizes were the largest for conscientiousness and test anxiety. Smaller effects sizes were found for emotional intelligence, the need for cognition. According to this M-A, general self-concept, emotional stability, extraversion, depression are practically independent of study success. The conclusion is that compared with other categories of student-related variable, personality variables show rather weak relations with study success.
-grades -retention -satisfaction	The authors find that personality traits have effects on academic achievement, but it depends on dimension of study success. For instance, Neuroticism has correlations with academic satisfaction, Conscientiousness is related to grades. Other three personality traits do not have significant relations to study success. have no significant impact on academic success."
online doctoral student GPA, self-reported, but verified via screen shot.	Conscientiousness relates to grades significantly. Grit does not add incremental value beyond other personality traits.

Summary of relevant findings/conclusions
Personality factors (noncognitive characteristics) are typically measured with self-assessments. Self-reports are coachable and fakable. There are numerous methods designed to minimize faking. However, more (large-scale) research is needed.

Type of personality assessment	Summary of relevant findings/conclusions
A list of 11 personal characteristics (critical thinking, internal motivation, ethical behavior, work ethic, persistent, reliable, attention to detail, self-reliance, organization, ability to work with Others, ability to Revise Self)	Admissions committees rated critical things the highest and ability to engage in self-revision was the lowest. See the Table in the study for all items.
A list of different personal qualities: all the big five personal factors and -Independence, leadership qualities, initiative/willingness to take a challenge, integrity, interests , etc.	Personality aspects did not appear amongst the most important selection criteria, according to admissions committees.
-Personality factors (the big five) -attitudinal factors (self-concept, self-efficacy, motivation, interests, social attitudes/values) -quasi-cognitive factors (creativity, emotional intelligence, cognitive style, metacognition)	Admissions committees draw attention to the importance of noncognitive factors along with cognitive ones. The authors also state that considering noncognitive factors has an advantage that gender and ethnic differences on noncognitive factors tend to be much smaller than for cognitive factors.

Authors & Year	Study type	OECD Country	Academic field	(Educational) level
MasterMind Europe (2017)	Q, cross-sectional	Europe	all disciplines	master's

Language Testing

Table A3.20 | *Language Testing. Evidence on Predictive Validity.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level	The language test examined
Abunawas (2014)	M-A	Across countries	across fields	undergraduate and graduate (International only)	TOEFL
Arcuino (2013)	Q, L.	US	technology, business, arts & sciences, education	master's (International only)	TOEFL iBT TOEFL PBT TOELF CBT IELTS
Bridgeman (2016)	R	n/a	across all disciplines, including STEM	master's and doctoral	TOEFL iBT
Burmeister et al. (2014)	Q, L	US	medical physics	master's and doctoral	TOEFL
Cho et al. (2012)	Q, L	US	across disciplines, including STEM	undergraduate and graduate	TOEFL iBT
Fu (2012)	Q, L	US	across different disciplines	master's and doctoral (international only)	TOEFL
Lee & Greene (2007)	Mixed-methods	US	science, technology, business, and humanities	graduate	The Computerized Enhanced ESL Placement Test [CEEPT]. In a large public university, where the study was conducted, CEEPT test is administered to all incoming international students whose TOEFL scores are below the campus or (if higher) departmental admission cutoff scores.

Type of personality assessment	Summary of relevant findings/conclusions
Personal skills	Personal skills are not used extensively for admissions to masters' programs (between a quarter and a third of programs). Personality also almost does not play a role in reasons for rejection.

Dimensions of study success examined	Summary of relevant findings/conclusions
GPA (both UGPA and GGPA)	TOEFL scores of international students were positively related to GPA. 47 effect sizes were examined, the overall effect size of .21 and 95% CI = .16 - .26. The educational level was not found as a significant moderator.
G-GPA	Relationship between TOEFL iBT and GGPA was significant, but weak and explained only 2% of variation both computer and paper versions of TOEFL did not have relationships to study success). There was no relationship between IELTS and G-GPA.
various, depending on a study included in the review	The author states that TOEFL essays and multiple-choice tests are nearly equally reliable. The overall conclusion is that essays of TOEFL should not been dismissed in selection because of unsupported claims of their questionable reliability or validity.
-faculty ratings -program Exit Exam -scores on the most difficult courses in the graduate program -First-year graduate GPA	TOEFL score was correlated to first-year GGPA, course average and faculty ratings (all p-values less than 0.05), but not exit exam.
GPA	The predictive value of TOEFL is small (graduate: r = .16; undergraduate: r = 18. It explained approx. 3% of the variance of GPA. Based on expectancy graphs, the authors conclude that students with higher TOEFL scores tend to earn higher GPAs. Moreover, TOEFL had incremental validity beyond other admissions requirements. The authors state that this relationship, though weak, is still meaningful.
first year GGPA	TOEFL had a significant correlation with first-year graduate GPA of international students. However, when undergraduate GPA and GRE-Verbal were already included in the model, TOEFL did not add to the prediction. When the author disentangled TOEFL on four skills, it appeared that only the Writing scale significantly predicted first-year GGPA. The author concludes that English writing skills was the most important predictor of first-year GPA among four language skills.
-first semester academic performance assessed via GPA -faculty evaluations, -student self-assessments	Based on quantitative analysis, ESL Placement test was not found as a significant predictor of study success. Mixed method analysis helped to explain this overall insignificant relationship. The authors find that the discipline plays a role. English was crucial to study success from the point of view of faculty members in humanities, but not that much in social sciences.

Authors & Year	Study type	OECD Country	Academic field	(Educational) level	The language test examined
Mathews (2007)	Mixed-method	Turkey	engineering (the majority of participants), political science, physics, management, educational psychology, biology, hospitality, tourism	graduate	A scale that takes into account the nature of the previous language use, for example, academic language use, intensive language study, and non-academic English use. It ranges from 1 (English on in high school to 7: English-medium during BA/MA in one of Turkish universities)
Sanford (2009)	Q, L, survey	US	engineering, computer science, physical sciences, social sciences, humanities, business, medicine, law, education, other	master's and doctoral (only international)	TOEFL
Wongtrirat (2010)	M-A	US	across disciplines	undergraduate and graduate levels (only international)	TOEFL
Zimmermann et al. (2018)	Q, L	Switzerland	STEM (the majority) & management, humanities, and social sciences (the minority)	master's	TOEFL iBT

Table A3.21 | *Language Testing. Evidence on Acceptability.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level
Ginther et al. (2014)	Mixed -methods	US & Australia	across disciplines	graduate
Lee & Greene (2007)	Mixed-methods	US	science, technology, business, and humanities	graduate

Dimensions of study success examined	Summary of relevant findings/conclusions
-success (completing a PhD degree abroad) -moderate success. (completing a master's degree without getting into a PhD program or running out of time, covered by a scholarship, during the PhD program). -failure (dropping out of graduate school without completing any degree)	Higher English proficiency relates positively to student success at foreign universities.
GPA time to degree completion	The coefficients for TOEFL were all non-significant. The author note that the sample lacked variability in TOEFL
GPA course completion	For graduate students: the TOEFL has small predictive ability toward GPAs of graduate level international students (.18). The TOEFL had almost no predictive ability (.085) on the course completion of international graduate students
GGPA	The TOEFL score explained 7%, additionally to UGPA and GRE scores. It is therefore concluded that TOEFL scores provide information above and beyond GRE scores.

The language test examined	Summary of relevant findings/conclusions
TOEFL IELTS PTE (the Pearson test of English)	Language assessments are often use just to determine whether applicants meet the institutional cut-off score. Beyond this application, language assessments had little impact on admissions decisions. However, the faculty members often expressed dissatisfaction with the English level of enrolled students. The authors bring attention to the fact that using the scores on language tests only as a cut-off, without understanding what the scores mean, creates misperceptions regarding the value of language tests. Most of admissions committees' members did not indicate a preference either for the TOEFL or the IELTS"
The Computerized Enhanced ESL Placement Test [CEEPT]. In a large public university, where the study was conducted, CEEPT test is administered to all incoming international students whose TOEFL scores are below the campus or (if higher) departmental admission cutoff scores.	Faculty members from different fields perceive the importance of proficiency in English differently: it is perceived less important in science disciplines than in humanities.

Authors & Year	Study type	OECD Country	Academic field	(Educational) level
MasterMind (2017)	Q, cross-sectional	Europe	all disciplines	master's
Mathews (2007)	Mixed-method	Turkey	engineering (most participants), political science, physics, management, educational psychology, biology, hospitality, tourism	graduate

Prior Research Experience

Table A3.22 | *Prior Research Experience. Evidence on Predictive Validity.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level	The aspects of prior research experience
Cox et al. (2009)	Q, L	US	computer science	doctoral	Whether or not a student wrote a thesis at the M.S. level
Gilmore et al. (2015)	Mixed-method	US	STEM	master's & doctoral	-presence of undergraduate research experience -duration of undergraduate research measured in semesters -degree of autonomy -collaboration/research network size - motivation The information gathered via interviews
Hall et al. (2017)	Q, L	US	biomedical	master's and doctoral	Amount of previous research experience, measured in months
Park et al. (2018)	Q, L	US	biomedical	doctoral	-Number of publications -prior research experience -research environment
Weiner et al. (2013)	Q, L	US	biology	doctoral	Years of research experience

The language test examined	Summary of relevant findings/conclusions
Language test/certificate (TOEFL, IELTS or other scores).	The results of the survey of MasterMind project indicate that approximately half of the master's programs in Europe require scores on language test during admissions (53.8%). At the same time, a bigger share of applicants (63.4%) indicate that they were requested to submit a language certificate. Some programs allow conditional admission (without a language certificate) and wait for the results. English and language related requirements are the third most reported reason unsuccessful admission (24.0%).
TOEFL	The qualitative findings of the study indicate that not all, but many consider that TOEFL scores do not reflect the proficiency in English. The test-takers state that receiving higher scores on the test does not mean that one can actually understand and speak English.

Dimensions of study success examined	Summary of relevant findings/conclusions
degree completion/dropout	Whether or not a student wrote a thesis at the M.S. level was a predictor of graduate study success. The authors hypothesize three possible reasons for this surprising, in their view, finding. (1) having experience with writing a master's thesis may give to a student some confidence about completing the Ph.D. project (so the student is less likely to drop out) (2) Having experience in writing a thesis makes students better prepared for the Ph.D. project (3) thesis-writing students are self-selected for success. (in that study, writing a master's thesis was an elective)
research skills performance measured by a research proposal, which was evaluated by a rubric for assessing scientific reasoning skills through writing.	Research experience during an undergraduate program was positively related to stronger research skills. The duration was the strongest predictor in level of research skill.
-time to degree -graduate student productivity as measured by several student publications	No correlation was found between amount of previous research experience and high or low productivity among admitted applicants. Importantly, in this study, all the applicants had a significant research experience, and they had no control population of accepted students with little-to no previous research experience.
-an introductory course of formal biomedical graduate training	The examined predictors did not have relation to performance on the introductory course.
ranking of students	Years of research experience was the strongest predictor of ranking.

Table A3.23 | *Prior Research Experience. Evidence on Acceptability.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level
Boyette-Davis (2011)	Q, cross-sectional	US	neuroscience	doctoral
Chari et al. (2019)	Q cross-sectional	US	physics	master's and doctoral
Kurysheva et al. (2019)	Q, cross-sectional	Netherlands	life and natural sciences	Master's

Various Selection Methods that do not Fall under Other Categories

Table A3.24 | *Various Selection Methods that do not Fall under Other Categories. Evidence on Predictive Validity.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level
Burmeister et. al. (2014)	Q, L	US	medical physics	master's and doctoral
Moneta-Koehler et al. (2017)	Q, L	US	biomedical	doctoral
Pacheco et al. (2015)	Q, L	US (Puerto-Rico)	biomedical	doctoral
Park et al. (2018)	Q, L	US	biomedical	doctoral
Sanford (2009)	Q, L (descriptive co-relational ex post facto study)	US	engineering, computer science, physical sciences, social sciences, humanities, business, medicine, law, education, other	master's and doctoral (only international)

Summary of relevant findings/conclusions
Having basic research experience was the most valued by the admissions committees. They valued less the publications and conference presentations
Publications and conference presentations were rated relatively low by masters' and PhD admissions committees. Prior research experience was rated as important (especially in doctoral admissions).
Research experience is valued in admissions to research masters' programs to a moderate degree. The criteria related to research background do not appear in top-10 most decisive factors. It might be explained by the fact, that the amount of research experience at bachelor level is very limited, so the admissions committees do not set high requirements for undergraduate students.

The selection method under examination	Dimensions of study success examined	Summary of relevant findings/ conclusions
Undergraduate degree type	-faculty ratings -course average -1st year GGPA, -performance on exit exam	An undergraduate degree (Physics) did not have a relationship with study success. The relationship was observed, however, between an undergraduate degree in engineering and faculty ratings. The authors indicate that the reason might be that those engineering students who enroll in their program, originally come from nuclear engineering. Nuclear engineering is relevant for their graduate studies in medical physics.
Undergraduate Institution Selectivity (UIS)	first semester grades	UIS was positively related to first semester grades (the lower the selectivity, the lower the first semester grades). A note: UIS was used as a control variable to test the incremental validity of the GRE.
A composite score (CS) that incorporates GPA, GRE, research experience, advanced course work or degrees, presentations, and publications	-degree completion -obtaining fellowships -time to degree	CS was predictive of degree completion. Moreover, it was positively related to obtaining fellowships and negatively to time to doctoral degree. Interestingly, other selection methods—GPA, science GPA, and GRE did not predict these three outcomes.
Competitiveness of previous research institution	-an introductory course of formal biomedical graduate training	a significant positive relationship was found between the predictor and the outcome.
The Non-Cognitive Questionnaire (NCQ). 8 scales: self-concept, self-appraisal, understanding and handling racism, the preference for long-term goals, the availability of a support person, leadership abilities, the community service, knowledge in the field outside of the formal experiences	-GGPA (self-reported) -time to degree completion	For GGPA: a few scales (self-concept; knowledge in the field) of NCQ were correlated to GGPA of masters' students. No scales were related to GGPA of doctoral students. For time to degree: The only NCQ scale which was related to time to degree was self-appraisal.

Authors & Year	Study type	OECD Country	Academic field	(Educational) level
van Os (2007)	Q, L	Netherlands	human movement science (FHMS), economics & business administration (FEBA), psychology & education (FPE) and social sciences (FSS)	master's
Willcockson et al. (2009)	Q, L	US	biomedical	master's and doctoral
Wollast et al. (2018)	Q, L	Belgium	all disciplines	doctoral
Zimmermann et al. (2015)	Q, L	Switzerland	computer sciences	master's

Table A3.25 | *Various Selection Methods that do not Fall under Other Categories. Evidence on Procedural Issues.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level
Posselt (2018)	Qualitative	US	not specified	graduate, participants – the selection committee members themselves.

The selection method under examination	Dimensions of study success examined	Summary of relevant findings/ conclusions
An assessment instrument for students with a vocational bachelor's degree wishing to pursue a master's program at the university. It focuses on the differences between higher vocational and academic education and makes them measurable.	-scores on exams for certain mathematical courses at those faculties - the total number of ECTS credits accumulated in the pre-master's program	The measure was a significant predictor. The extent of prediction varies across faculties.
Prior graduate degree, usually in a health -related discipline such as nursing.	-performance on an introductory course of formal biomedical graduate training	Prior graduate degree was positively related to Mastery and negatively—to Failure. The authors explain this by the fact that students with a graduate degree had already obtained certain skills like time management etc.
-Research field (1) humanities, (2) social sciences, (3) health sciences, and (4) science and technology -Same university for the undergraduate degree and for the doctoral degree -Same field for the undergraduate degree and for the doctoral degree	degree completion/ dropout	When considered separately: -PhD students in sciences and technology have higher odds of completing their project than students in other disciplines. -doing a PhD project in the same university where undergraduate degree was done, is related to higher completion rates -the same applies to the field of study (on a marginal level) When considered in one model together and with other predictors, research fields stayed significant and other two did not
-Rate of progress: the number of credits obtained in the bachelor's program divided by its duration -Duration: the time required to finish the entire bachelor's program	GGPA	Rate of progress and duration were not shown as strong predictors of GGPA.

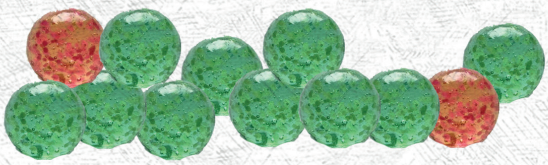
Procedural Issues	Summary of relevant findings/conclusions
Academic pedigree	Many HEIs claim that they wish to reduce racial and SES inequalities. However, they proceed with practicing academic pedigree, which promotes these phenomena.

Table A3.26 | *Various Selection Methods that do not Fall under Other Categories. Evidence on Acceptability.*

Authors & Year	Study type	OECD Country	Academic field	(Educational) level
Boyette-Davis (2011)	Q, cross-sectional	US	neuroscience	doctoral
Chari et al. (2019)	Q cross-sectional	US	physics	master's and doctoral
MasterMind Europe (2017)	Q, cross-sectional	Europe	all disciplines	master's
Kurysheva et al. (2019)	Q, cross-sectional	Netherlands	life and natural sciences	master's
van Os (2007)	Q, cross-sectional	Netherlands	human movement science, economics & business administration, psychology & education, and social sciences.	master's

The selection method under examination	Summary of relevant findings/conclusions
-Extracurricular activities -Teaching experience	-extracurricular activities (mean = 3.45) -teaching experience (mean=2.61). The scale was from 1 (not important at all) till 7 (very important) The authors find that these two selection criteria/methods are not of high importance in admission decisions. Also, none of admissions' committees members indicated that these are important skills.
Undergraduate courses taken (UCT)	UCT was an important criterion for both masters' and doctoral admissions. The authors note that some students do not have an opportunity to have "canonical" courses in their undergraduate programs which might disadvantage them in the admissions process to these masters' programs.
Quantitative skills Work experience CV Essay Photograph	Quantitative skills were not very important in admission decisions: around 9% of program indicated that this criterion must be met to be eligible for a master's program and around 25% of programs indicated that this will be an advantage. Work experience: almost no programs require it (7%), but for almost a half (46%), it would be an advantage. CV: more than a half of programs require it (58%), but only for one fifth of the programs, the information presented there give an advantage for the admission decision. Essay is almost never required (only in 6% programs) and considered as an advantage in 14% cases Photograph: is required almost by third of the programs (30%) and representatives of a few programs indicated that it gives an advantage in admission process (5%).
-Type of prior academic background -Type of prior education institution -Number of credit points earned for relevant courses -Preliminary plans for the master's program -Time management skills -Understanding of social relevance of research -Integrity, fairness/honesty	-Type of prior academic background and type of prior education institution are in top-10 of most important selection criteria both at the life and natural sciences' programs. - number of credit points earned for relevant courses is also in top 10 at the natural sciences -preliminary plans for the master's program, ambition for a chosen master program is in top-10 of criteria at the life sciences. -Time management skills, understanding of social relevance of research, integrity and fairness/honesty are the least used as selection methods by selection committees of research master programs in the life and natural sciences.
An assessment instrument for students with a vocational bachelor's degree wishing to pursue a master's program at the university. It focuses on the differences between higher vocational and academic education and makes them measurable.	By students: The students perceived English Reading Comprehension and Mathematics & Statistics subtests the most relevant. Scientific Reasoning also scored high, wherever applicable. Overall, more than 50% gave the level a satisfactory rating to this assessment. By counselors: The counsellors rated the assessment positively, also for selection purposes. However, there were substantial differences between faculties.

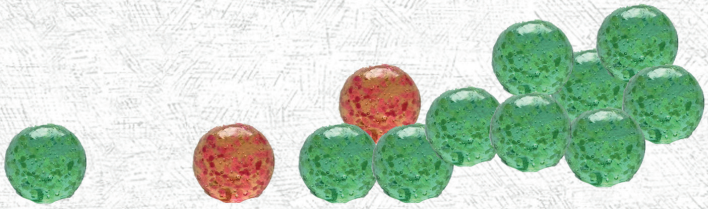
3



Acknowledgements

The authors appreciate the assistance of Anne-Petra Rozendal, Rianne Bouwmeester, Elianne Gerrits, Lieselotte Postmes, Shirrinka Goubitz in coding admissions criteria into categories. We are also thankful to Farah Kools, Els van der Vlist, Rianne Bouwmeester, Sanne de Jong, and Kirsten Boersma-van Nierop for checking the text of the surveys and their useful suggestions. We thank Twan Maintz, Pauline Krijgsheld, Elena Donner, Yolande Lutteke, and Shirrinka Goubitz for their help in informing the participants about this study. We also appreciate the help of Cristina Spoiala and Christine Fox with copy editing.

How do Admissions Committees Select? Do Applicants Know How They Are Being Selected? Selection Criteria and Transparency at a Dutch University



Author note

An adapted version of this chapter is published as "Kurysheva A., van Rijen H. V. M., & Dilaver G. (2019). How do admissions committees select? Do applicants know how they select? Selection criteria and transparency at a Dutch University. Tertiary Education and Management, 25, 367–388. https://doi.org/10.1007/s11233-019-09050-z"

Kurysheva, H. V. M. van Rijen, and G. Dilaver conceptualized the overarching research aims. A. Kurysheva developed the methodology, designed the survey, conducted formal analysis, prepared the original draft. H. V. M. van Rijen and G. Dilaver provided advice on communication strategies with participants and organization of data collection as well as conducted critical review of the manuscript.

Abstract

In this study, we investigated the currently applied selective admissions criteria and methods of two-year research masters' programs in the Graduate Schools of Life Sciences and Natural Sciences of Utrecht University (the Netherlands). In addition, we also evaluated how transparent these selection criteria are to applicants. Both admissions staff members and applicants participated. To determine the admissions criteria that are important for admissions decisions, we first ranked 51 different admissions criteria. We then categorized them into six domains: academic background, grades, cognitive ability, research background, personality and personal competencies, and motivation factors. To evaluate transparency, we contrasted the perceptions of applicants with the actual importance of admissions criteria, as reported by admissions staff members. We found that admissions criteria related to the domain of personality and personal competencies are less important in admissions decisions than criteria related to grades, academic background, and motivation. Applicants found the admissions decisions transparent to a moderate degree. This study also revealed that admissions committees use selection criteria and methods both with and without predictive value for graduate study success. Moreover, some of the currently applied admissions methods can be prone to admissions biases. Based on our findings, we strongly recommend admissions committees to use selection criteria and methods that are evidence-based, resistant to admissions biases, and transparent to the applicants.

Keywords: admissions, selection, graduates, applicants, transparency

3.1 Introduction

In the Netherlands, the number of masters' students attending research universities is steadily increasing (Association of Universities in the Netherlands, 2018; see Fig. 1). This increase is partly associated with the growing number of international students, which has nearly tripled in the last ten years (Fig. 1). The international attractiveness of Dutch graduate education is not surprising, since all 13 Dutch research universities appear in the top 250 world university rankings (Times Higher Education, 2018) and offer numerous English-taught programs, especially at the master's level. Therefore, selective admissions are becoming both necessary and complex due to the growing numbers of applicants and increased diversification of the application files. Meanwhile, two recent legal changes in the Dutch admissions university system—(1) a switch from open admissions and lottery systems to selective admissions and (2) the requirement to fully complete a bachelor's degree before the start of a master's program—have contributed to societal expectations of fair, transparent, and inclusive higher education selective admissions.

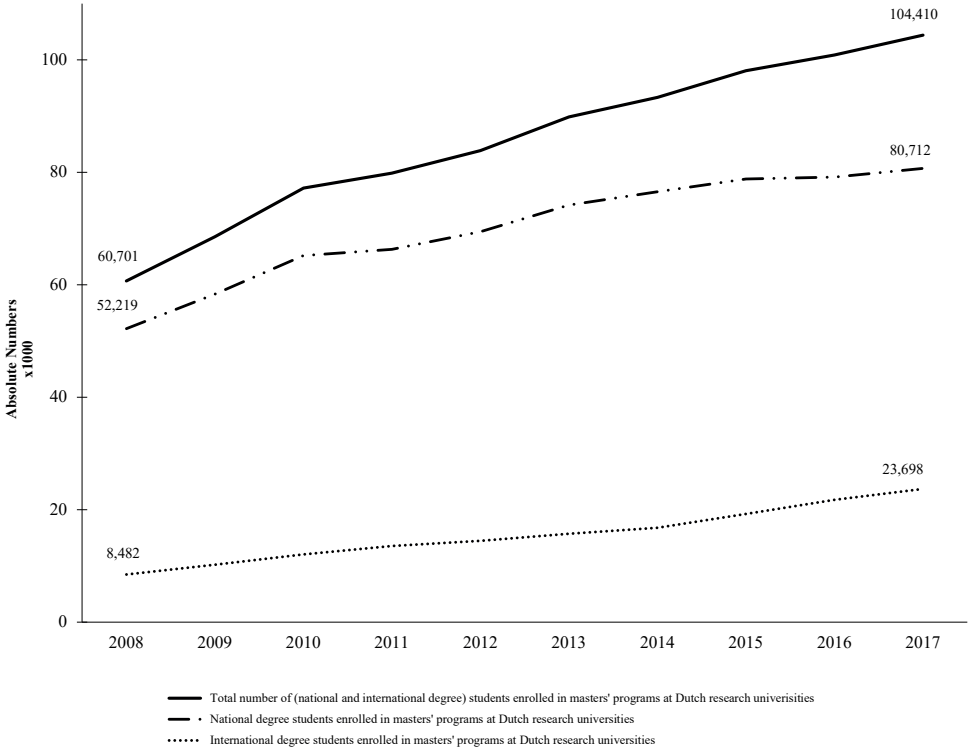


Figure 1 | *Numbers of Masters' Students at Dutch Research Universities during the Period 2008–2017.*
Note. Statistics of the Association of research universities in the Netherlands (Association of Universities in the Netherlands [VSNU], 2018).

3.1.1 Evaluative Quality Principles for Selection Procedures

Patterson and Ferguson (2010) suggested twelve evaluative quality principles for selection procedures that should be considered when designing a selection process (Table 1). They draw specific attention to the criteria of *validity, reliability, fairness*, and *applicant reactions* as crucially important in the design of selective admissions. Validity refers to how accurate a selection method predicts future outcomes. Reliability refers to consistency of selective methods across different conditions (Cleland et al., 2012). Fairness, in broad terms, refers to absence of biases toward different applicant groups, for example, minorities (Tillema et al., 2011). Applicant reactions refer to perceived fairness and clarity of the selective admissions process. Closely related both to fairness and to applicant reactions is transparency that refers to an applicants' assessment of the selective admissions process rather than its content: It indicates the necessity of providing clear and understandable requirements (Tillema et al., 2011). Transparent admissions allow applicants to estimate whether their academic background, prior achievements, and noncognitive attributes are suitable for a program, and what they could learn, improve, or provide in order to be selected.

Table 1| *Evaluative Quality Principles for Student Selection.*

Reliability and validity of selection methods
Candidate / applicant reactions
Ease of interpretation
Generality of use
Cost
Practicality / administrative convenience
Legality
Availability of analytical expertise
Fairness
Educational impact
Mechanisms for obtaining feedback
Arrangements for ongoing evaluation and development

Note. Adapted from "Selection for medical education and training" by F. Patterson and E. Ferguson, 2010, In T. Swanwick (Ed.), *Understanding Medical Education: Evidence, Theory and Practice*, p. 356. https://doi.org/10.1002/9781444320282.ch24

Disregarding these fundamental evaluation quality principles in a selective admissions process can have a wide range of negative consequences. Namely, using invalid and unreliable selective admissions methods can lead to a higher number of false positive cases: students who are admitted to a program but are unable to complete it (Admissions to Higher Education Steering Group, 2004), and a higher number of false negative cases–rejected applicants who would have been able to successfully complete

the program if they had been selected (Ooijen-van der Linden et al., 2017). Insufficient transparency or misleading information puts applicants at a disadvantage: They may fail to acquire knowledge and skills that are important for a graduate program or to realistically assess the likelihood to be accepted into a graduate program (Nauta, 2000). Insufficient transparency may also discourage applicants with low self-confidence from applying (Admissions to Higher Education Steering Group, 2004).

A call for more transparency by decentralized selection has been made by policy makers (Adnett et al., 2011; Admissions to Higher Education Steering Group, 2004), higher education institutions (HEI; e.g., Utrecht University, 2017), and by the students themselves (The Dutch National Students Association, 2014). Following this call, some universities began publishing minimum, median, and top study scores (Bagshaw, 2016), standardized test scores, and background information of their admitted students such as race, ethnicity, and geographical origins (La Noue, 2003). While these actions have helped create more transparency regarding required admissions documents and diversity of admitted students, what remains ambiguous is how admissions committees actually select applicants (La Noue, 2003).

3.1.2 How Do Admissions Committees Select?

In the context of graduate education, where the admissions process is often decentralized, selective admissions decisions are made by academic staff members who participate in a program's committee (referred to as admissions committee members). Research across different fields has shown that admissions committee members look for indicators on both academic background and performance but also on personality, motivation, and attitude.

For academic background and performance, admissions committee members consider important a match or fit between interests and skills of an applicant and what the program offers (Karazsia & McMurtry, 2012; Karazsia et al., 2013; Posselt, 2016; Walpole et al., 2002); the quality of applicants' undergraduate experience such as prior institution–where and how an applicant received the undergraduate credits (Kogan et al., 2015; Walpole et al., 2002); prior degree in a specific field, transcripts, an applicant's work experience, and grade point average (GPA; MasterMind Europe, 2017). Regarding criteria related to personality, motivation, and attitudes, admissions committee members place value on an applicant's integrity and honesty (Walpole et al., 2002), communication skills, critical thinking, ethical behavior (Conlon et al., 2012), and personal skills (MasterMind, 2017).

The measurement of some of these criteria have been criticized for being prone to biases in selective admissions practice (Kira Talent, 2018; Posselt, 2016). Namely, the assessment of personality, motivation, and attitudes which can be found in interviews, recommendation letters, curriculum vitas (CVs), and motivation letters. Interviews, CVs, and motivation letters represent self-reported measures. When applied in a high-stake situation of selective admissions, they are susceptible to faking by the applicants

(Niessen et al., 2017). When interpreting recommendation letters, a measure which is not self-reported, the admissions committee members sometimes indicate a need "to read between the lines" or "to detect an extra glow of enthusiasm about the student" (Walpole et al., 2002, p. 5), which indicates the susceptibility of this selection method to biases.

3.1.3 The Current Study

In this study, we consider two overarching questions: How do admissions committees select and do applicants know how they select? For this study, we first examine the current selection criteria and methods applied in practice for research-oriented programs at two large European graduate schools. We then look at the transparency of that criteria toward applicants. Additionally, we examine which selection methods are used for assessment of these criteria. The aim of this study is to determine the selective admissions criteria that are important for decision-making for research masters' programs and to investigate whether they are accurately perceived by the applicants.

3.1.3.1 Context of the study

Utrecht University is a HEI that offers research-oriented education in contrast to higher professional education which is offered by universities of applied sciences. A specific focus on research-oriented education means that the main goal of study programs is "independent academic participation or the professional use of academic knowledge" (Nuffic, 2015, p. 11). For example, the mission of one of the graduate schools participating in this study is "to train future scientists who will be capable of drawing on their acquired knowledge, understanding, and skills to conduct top-class research around the world" in academia or industry (Graduate School of Life Sciences, n.d.). During a two-year master's program, a student focuses on a specific field and learns about research methods. A final thesis is an obligatory component of a research-oriented study program.

Two Graduate Schools from Utrecht University were involved in this research: the Graduate School of Life Sciences (GSLS) and the Graduate School of Natural Sciences (GSNS). The GSLS is organized by three faculties: Medicine, Veterinary Sciences, and Science (biology, pharmaceutical sciences, and chemistry). The GSNS is organized by the faculty of Science (mathematics, physics, and information and computing sciences). At the GSLS, there were over 2000 applicants in 2018, however, there was only room for 450 students. This limited capacity has created an urgency to evaluate the applied selective admissions criteria and methods in terms of their compliance with evaluative quality principles as described in the literature (Patterson & Ferguson, 2010) and requested by the university itself (Utrecht University Education Guideline, 2017).

3.1.3.2 Research Questions

The curriculum of the life sciences research masters' programs at the GSLS is internship-focused, while the curriculum of the natural sciences research masters' programs at the

GSNS is course-focused. We would expect that the relative importance of admissions criteria is different for different types of curricula. Therefore, our first research question is*: What kind of selection criteria are reported to be important in admissions decision-making at the Life and Natural Sciences graduate schools by the designated admissions staff?*

Not only the criteria themselves, but also their measurements should meet the evaluation quality principles. We, therefore, examine how certain criteria are measured. This brings us to the second research question: *What selection methods are most frequently applied?* In addition, we aim to empirically analyze transparency of admissions and, therefore, our third research question is: *Are the applied criteria transparent for the applicants?* These questions are tackled in a cross-sectional descriptive study that uses a survey methodology. We also discuss validity, reliability, and fairness of applied selection criteria and methods based on prior literature in the discussion session.

3.2 Method

3.2.1 Procedure and Measures

The study was approved by the Netherlands Association for Medical Education Ethical Review Board (dossier number: 969). An invitation to participate in the survey, containing the URL for the online survey, an information letter, and a consent form were sent to the participants by email.

3.2.2 Admissions Criteria

The 53 possible admissions criteria were derived from the graduate schools' admissions webpages and admissions guidelines, scientific literature on criteria that are valued by selection committees worldwide (Conlon et al., 2012; Kogan et al., 2015; MasterMind Europe, 2017; Walpole et al., 2002; Weiner, 2014), scientific literature on evidence-based determinants of study success (Craig et al., 2004; Hall et al., 2017; Kuncel et al., 2001, 2004; Pacheco et al., 2015; Poropat, 2009; Trapmann et al., 2007; von Stumm et al., 2011), and exploratory conversations with the admissions committee members on what they look for in admissions applications. In the survey, rating of each 53 admissions criteria was on a 5-point Likert scale (from 1 = not very important to 5 = extremely important) and 0 indicated that the criterion was not used by admissions committee members/was not perceived by applicants as being used. As two of the suggested criteria—"topic of bachelor's thesis/research project/internship" and "grade for bachelor's thesis/research project(s)"—were often found to be not available at the moment when students apply, we excluded them from the analysis, which left us with a list of 51 admissions criteria.

Six coders—researchers in the higher education field and experts in (bio)medical education—were asked to sort the 51 admissions criteria into categories (general

admissions domains). The names of some domains were suggested to them based on what selection methods primarily measure (e.g., motivation, cognitive ability, etc.), but the experts were free to add/change the domains if they saw a possibility of a better fitting categorization. After their assessment, six final categories emerged. The confirmatory factor analysis with Oblimin rotation delivered a similar but not completely identical solution. We explain this by the small sample size of the program coordinators and relatively small sample size of natural science applicants (for a reliable factor analysis, there should be at least four or five times as many observations as there are criteria). Therefore, we report the results based on the experts' categorization: "academic background", "grades", "cognitive ability", "research background", "personality and personal competencies", and "motivation factors" (the average percentage of agreement was 66%). Their interrater reliabilities (Cronbach's alphas) are presented in Table 2. The coders could not assign two criteria, "understanding of social relevance of research" and "qualitative assessments by previous mentors". Therefore, these two criteria were excluded from the univariate analysis on domains.

Table 2 | *Admissions Domains: Cronbach's Alphas and Selective Admissions Criteria Included per Domain.*

Domain	The GSLS admissions committees members	The GSLS applicants	The GSNS admissions committees members	The GSNS applicants	Specific criteria included
Academic background (11 criteria)	.51	.66	.70	.72	-Type of prior education institution (a preference for either a Dutch research university/comparable international university, or for university college, or university of applied sciences (HBO) -Academic writing skills -Oral and written proficiency in English at an advance level (based on standardized language tests; applicable for international students only) -Oral and written proficiency in English at an advance level based on additional language assessment (based on interviews/documents) -Experience in teamwork -Correct communication both in writing and orally -Prior academic background (Biomedical sciences, Biological sciences, etc.) -Content of previous education matches with master's program -Number of credit points earned for relevant course(s) -Number of EC (credit points) earned during prior education -Time to prior degree (i.e., time between starting date and graduation date of a bachelor program)
Grades (4 criteria)	.87	.65	.61	.63	-Undergraduate Grade Point Average (UGPA) -Increase in grades across the years of prior education -Cum Laude graduation -Grades for relevant courses

Domain	The GSLS admissions committees members	The GSLS applicants	The GSNS admissions committees members	The GSNS applicants	Specific criteria included
Cognitive ability (6 criteria)	.89	.77	.87	.79	-General cognitive ability (Intelligence) -Intellectual curiosity (Intellectual engagement in the pursuit of knowledge) -Critical thinking, skills, logic, problem-solving skills -Participation in honors programs -The ability to understand the scientific method -Talent
Research background (5 criteria)	.73	.80	.54	.83	-Previous research experience (amount and quality) -A clear interest in a multidisciplinary research approach -Quality of research experience -The ability to design and carry out research projects -Mastery of field relevant methods and techniques
Personality and personal competencies (14 criteria)	.94	.94	.96	.95	-Agreeableness (likability and friendliness) -Conscientiousness (trustworthiness and will to achieve) -Emotional stability (emotional adjustment as the opposite to anxiety and neuroticism) -Extraversion (activity and sociability) -Openness to experience (imaginativeness, broadmindedness, and artistic sensibility) -Initiative/willingness to take a challenge -Interpersonal skills (the ability to communicate or interact well with other people) -Interests (sports, creative activities etc.) -Integrity, fairness/honesty -Maturity/responsibility/work habits -Leadership qualities -Independence -Volunteering experience (amount and quality) -Time management skills/Timeliness
Motivation factors (9 criteria)	.89	.82	.87	.88	-Knowledge about program applying for -Motivation for studying at UU -Motivation for doing research -Motivation for doing a master's program -Motivation for this particular program -The desire to acquire certain skills relevant for this program (e.g., science, laboratory-, business skills) -Preliminary plans for the master's program (e.g., having an idea about the laboratory where a student wants to have an internship) -Ambitions: what a student wants to achieve with the concrete master's program? -Travelling experience

Note. Cronbach's alpha was used for reliability estimate of the six domains. Two criteria are not included into these domains due to no agreement between the experts: "Qualitative assessments by previous mentors (recommendation letter writer references/ratings)" and "Understanding of social relevance of research".

Table 3 | *Participants' Characteristics and Response Rate.*

	GSLS	GSNS
Admissions committee members		
Response rate admissions committee members (*N* responded admissions committee members)	33% (*N* = 16)	100% (*N* = 10)
Coverage of programs (*N* responded masters' programs)	80% (*N* = 12)	100% (*N* = 10)
Cohorts of applicants 2017–2018 & 2018–2019		
Response rate (*N* responded applicants)	22% (*N* = 327)	13% (*N* = 153)
Coverage of programs	100% (*N* = 15)	100% (*N* = 10)
% accepted applicants out of all respondents	89%[a]	94%
Rejected applicants	11%	6%
Bachelor's degree at Utrecht University	42% (n = 138)	26% (n = 39)
Bachelor's degree at another Dutch research university	21% (n = 68)	26% (n = 39)
Bachelor's degree at a Dutch university college	5% (n = 16)	0
Bachelor's degree at a Dutch university of applied sciences	4% (n = 12)	9% (n = 14)
Bachelor's degree at a college or a university outside of the Netherlands	26% (n = 86)	39% (n = 59)
Other type of prior education (e.g., both bachelor's and master's degrees)	2% (n = 7)	0.01% (n = 2)

Note. [a] Nine students responded who were still on the waiting list. However, their data is excluded from the analysis since they have not undergone the whole selective admissions process at that moment.

3.2.3 Admissions Methods (Documents and Tools)[11]

The admissions committee members were asked to indicate which methods (i.e., documents and tools such as transcripts/grade lists, course content, CV, motivation letter, letter(s) of recommendation, interview, English language test, writing sample, personal acquaintance) they use in assessing different groups of admission criteria (such as education background, disciplinary background, cognitive ability, personality, motivation, and attitude).

3.2.4 Participants and Response

At each Graduate School, admissions committee members and applicants to two-year research masters' programs (2017–2018 and 2018–2019 cohorts) participated. The participants' characteristics and response rate are presented in the Table 3. Relatively low response rate was associated with rejected applicants not being inclined to respond to the survey from a graduate school that rejected their applications.

11 Paragraphs 3.2.3 and 3.2.4 were missing from the version of thesis, sent to the Assessment Committee. They are present in the published version of this article (2019). Therefore, the mistake is corrected and they are added to the thesis after proofreading the text.

3.2.5 Data Analysis

Analysis was conducted in SPSS 25. There were no missing values in the responses of the admissions committee members from both graduate schools. A few values were missing in the responses of applicants: 0.2% of values in responses of the GSLS applicants and 0.4% of values in responses of the GSNS applicants. To handle these missing values, the expectation-maximization (EM) algorithm in SPSS was used. In analysis, first, the descriptive statistics were obtained to rank the importance of admissions criteria, and the frequencies of selection methods usage were obtained. To assess transparency, the importance of admissions criteria, as reported by admissions committee members, was contrasted to the importance as perceived by applicants. This was done using Spearman correlation and 2-factorial analysis of variance (ANOVA; 1st factor = admissions committee members versus applicants, 2nd factor = a domain of admissions criteria).

3.3 Results

3.3.1 Research Question #1

We examined the importance of suggested admissions criteria in selective admissions decisions on two levels: on the level of specific selection criteria and on the level of general admissions domains of those criteria. For examination of specific admissions criteria's importance, the means of 51 selective admissions criteria were ranked according to their importance in actual admissions decisions as reported by admissions committee members (Table 4).

3.3.1.1 On the Level of Specific Criteria

For the GSLS, the most important criteria were related to the domains "motivation factors" (n = 5), "academic background" (n = 3), and "grades" (n = 2). No criteria specifically related to research background, cognitive ability or personality aspects appeared among the most important. For the GSNS, the highest positions were given to criteria related to "academic background" (n = 5) and "grades" (n = 2). There was also one criterion ("mastery of field relevant methods and techniques") related to research background (#6). In contrast to the GSLS, specific criteria related to motivation also appeared in the top-10, but only two of them and not in the highest positions (#6 and #10). Despite the difference in ranking per criterion, the correlational analysis of the whole list of admissions criteria revealed that overall, the importance which both the GSLS and the GSNS members of admissions committees assigned to selective admissions criteria was similar to a high extent: $r_{Pearson}(49) = .76$, $p < .001$.

Table 4 | *Ranks, Means and Standard Deviations of Selective Admissions Criteria: Admissions Committee Members Perspectives and Students Perceptions at the Graduate Schools of Life Sciences and Natural Sciences.*

Graduate School of Life Sciences				
Admissions criterion	**Admissions committees members**		**Applicants**	
	Rank	***M (SD)***	**Rank**	***M (SD)***
Motivation for this particular program	1	4.44 (0.81)	1	4.34 (0.90)
Content of previous education matches with master's program	2	4.31 (0.79)	7	3.73 (0.93)
Type of prior educational institution (a preference for either a Dutch research university/comparable international university, or for university college, or university of applied sciences (HBO)	3	4.00 (1.03)	16	3.47 (1.16)
Motivation for doing research	4	3.94 (1.18)	2	4.00 (0.91)
Prior academic background (Biomedical sciences. Biological sciences. etc.)	5	3.81 (1.05)	4.5	3.86 (0.83)
Knowledge about the program applying for	6	3.63 (1.09)	8	3.73 (0.98)
Preliminary plans for the master's program (e.g., having an idea about the laboratory where the student wants to have an internship)	7	3.56 (1.50)	36.5	2.86 (1.32)
Undergraduate Grade Point Average (UGPA)	8.5	3.44 (1.26)	24	3.26 (1.04)
Ambitions: what a student wants to achieve with the concrete master's program?	8.5	3.44 (1.41)	13	3.54 (1.05)
Motivation for doing a master's program	10.5	3.38 (1.50)	4.5	3.86 (1.03)
Grades for relevant courses	10.5	3.38 (1.02)	17	3.44 (1.05)
Previous research experience (amount and quality)	12	3.19 (1.52)	20	3.37 (1.26)
Oral and written proficiency in English at an advanced level (based on standardized language tests; applicable for international students only)	13.5	3.06 (1.69)	41	2.56 (1.93)
A clear interest in a multidisciplinary research approach	13.5	3.06 (1.61)	12	3.56 (1.17)
The desire to acquire certain skills relevant for the program (e.g., science-. laboratory- or business- skills)	15.5	3.00 (1.59)	9	3.73 (1.08)
Number of credit points earned for relevant course(s)	15.5	3.00 (1.37)	28	3.17 (1.19)
Intellectual curiosity (Intellectual engagement in the pursuit of knowledge)	17	2.94 (1.95)	3	3.96 (1.08)
Correct communication both in writing and orally	19.5	2.81 (1.80)	11	3.63 (1.14)
Oral and written proficiency in English at an advanced level based on additional language assessment (based on interviews/ documents)	19.5	2.81 (1.80)	31	3.09 (1.44)
Qualitative assessments by previous mentors (Recommendation letter writer references/ratings)	19.5	2.81 (1.38)	36.5	2.86 (1.51)
Participation in honors programs	19.5	2.81 (1.47)	40	2.58 (1.32)
Motivation for studying at UU	23.5	2.69 (1.58)	14	3.52 (1.13)
Mastery of field relevant methods and techniques	23.5	2.69 (1.49)	34	3.01 (1.14)
Cum Laude graduation	23.5	2.69 (1.74)	45	2.32 (1.44)
Increase in grades across the years of prior education	23.5	2.69 (1.45)	39	2.60 (1.31)
Quality of research experience	26	2.63 (1.78)	30	3.12 (1.27)

Graduate School of Natural Sciences				
Admissions criterion	**Admissions committees members**		**Applicants**	
	Rank	***M (SD)***	**Rank**	***M (SD)***
Content of previous education matches with master's program	1	4.40 (0.70)	3	3.87 (1.10)
Prior academic background (Astronomy, Artificial Intelligence, Physics, etc.)	2	4.10 (0.88)	6	3.76 (0.94)
Type of prior educational institution (a preference for either a Dutch research university/comparable international university, or for university college, or university of applied sciences (HBO)	3	3.80 (0.92)	8	3.66 (1.18)
Number of credit points earned for relevant course(s)	4	3.70 (0.67)	16	3.30 (1.32)
Grades for relevant courses	6	3.60 (0.52)	18	3.26 (1.14)
Mastery of field relevant methods and techniques	6	3.60 (1.51)	20	3.16 (1.22)
Motivation for this particular program	6	3.60 (1.51)	1	4.20 (0.89)
Undergraduate Grade Point Average (UGPA)	8	3.50 (1.18)	24	3.03 (1.12)
Oral and written proficiency in English at an advanced level (based on standardized language tests; applicable for international students only)	9	3.40 (1.51)	34	2.73 (1.82)
Motivation for doing research	10	3.30 (0.82)	10	3.49 (1.16)
Number of credit points earned during prior education	11.5	3.20 (1.40)	22	3.10 (1.30)
Knowledge about the program applying for	11.5	3.20 (0.92)	2	3.90 (0.84)
Qualitative assessments by previous mentors (Recommendation letter writer references/ratings)	13	3.00 (1.15)	40	2.42 (1.61)
Talent	14.5	2.70 (2.06)	41	2.36 (1.44)
Initiative / willingness to take a challenge	14.5	2.70 (1.49)	9	3.58 (1.26)
Cum Laude graduation	17	2.60 (2.01)	44	2.15 (1.47)
Motivation for studying at UU	17	2.60 (1.07)	12	3.44 (1.13)
Time to prior degree (i.e., time between starting date and graduation date of a bachelor program)	17	2.60 (0.84)	39	2.42 (1.22)
The desire to acquire certain skills relevant for the program (e.g., science-. laboratory- or business- skills)	21	2.50 (1.51)	14	3.37 (1.33)
Motivation for doing a master's program	21	2.50 (1.65)	4	3.82 (1.02)
The ability to understand the scientific method	21	2.50 (2.01)	13	3.39 (1.30)
Previous research experience (amount and quality)	21	2.50 (1.35)	35	2.69 (1.31)
Increase in grades across the years of prior education	21	2.50 (1.27)	42	2.28 (1.27)
Oral and written proficiency in English at an advanced level based on additional language assessment (based on interviews/ documents)	24	2.40 (1.90)	38	2.60 (1.69)
Ambitions: what a student wants to achieve with the concrete master's program?	25.5	2.30 (1.16)	11	3.48 (1.18)
Academic writings skills	25.5	2.30 (1.95)	26	3.01 (1.34)
Correct communication both in writing and orally	27	2.20 (1.69)	15	3.33 (1.29)

Graduate School of Life Sciences				
Admissions criterion	**Admissions committees members**		**Applicants**	
	Rank	***M (SD)***	**Rank**	***M (SD)***
Time to prior degree (i.e., time between starting date and graduation date of a bachelor program)	27	2.50 (1.21)	44	2.32 (1.16)
Initiative / willingness to take a challenge	28.5	2.31 (1.99)	10	3.68 (1.08)
The ability to understand the scientific method	28.5	2.31 (1.82)	15	3.51 (1.30)
Conscientiousness (trustworthiness and will to achieve)	30	2.19 (2.04)	18	3.40 (1.28)
Interpersonal skills (the ability to communicate or interact well with other people)	32	2.13 (1.82)	26	3.18 (1.33)
Traveling experience	32	2.13 (1.41)	51	1.70 (1.09)
General cognitive ability (intelligence)	32	2.13 (2.00)	23	3.26 (1.16)
Interests (sports, creative activities, etc.)	34	2.06 (1.34)	49	2.18 (1.19)
Academic writing skills	36	2.00 (1.93)	25	3.24 (1.23)
Volunteering experience (amount and quality)	36	2.00 (1.41)	50	2.08 (1.15)
Extraversion (activity and sociability)	36	2.00 (1.41)	48	2.23 (1.31)
Independence	38.5	1.94 (1.84)	26	3.19 (1.48)
Critical thinking skills, logic, problem-solving skills	38.5	1.94 (2.05)	6	3.80 (1.18)
Understanding of social relevance of research	40.5	1.88 (1.54)	33	3.01 (1.38)
Number of EC (credit points) earned during prior education	40.5	1.88 (1.36)	38	2.86 (1.31)
Talent	42.5	1.81 (1.94)	42	2.51 (1.36)
Maturity/responsibility/work habits	42.5	1.81 (1.72)	21	3.34 (1.43)
The ability to design and carry out research projects	44.5	1.75 (1.73)	29	3.17 (1.29)
Agreeableness (likability and friendliness)	44.5	1.63 (1.63)	43	2.45 (1.33)
Experience in teamwork	46	1.56 (1.55)	19	3.37 (1.20)
Openness to experience (imaginativeness, broad-mindedness, and artistic sensibility)	47	1.44 (1.71)	35	2.98 (1.38)
Emotional stability	48	1.38 (1.59)	46	2.28 (1.49)
Integrity, fairness/honesty	49	1.31 (1.82)	22	3.32 (1.57)
Time management skills / Timeliness	50	1.25 (1.73)	32	3.01 (1.42)
Leadership qualities	51	1.13 (1.54)	47	2.26 (1.32)

3.3.1.2 On the Level of General Admissions Domains

After examination of specific selection criteria, we turned to a higher order level: exploration of relative importance of general selective admissions domains based on these criteria. Their means and standard deviations are presented in Table 5.

For the GSLS, a univariate ANOVA revealed that there was a significant effect of domain on average importance of criteria; the effect was of a large effect size, $F(5, 43) = 9.02$, $p < .001$, $\eta^2 = .51$. Hochberg's post-hoc examination indicated that the "personality and personal competencies" domain was less important compared to three other domains:

Graduate School of Natural Sciences				
Admissions criterion	**Admissions committees members**		**Applicants**	
	Rank	***M (SD)***	**Rank**	***M (SD)***
Intellectual curiosity (Intellectual engagement in the pursuit of knowledge)	28	2.10 (2.18)	5	3.79 (1.19)
Critical thinking skills, logic, problem-solving skills	29.5	2.00 (2.16)	7	3.67 (1.25)
The ability to design and carry out research projects	29.5	2.00 (1.41)	28	2.84 (1.49)
A clear interest in a multidisciplinary research approach	32	1.90 (1.91)	19	3.19 (1.37)
Participation in honors programs	32	1.90 (1.66)	43	2.20 (1.37)
Quality of research experience	32	1.90 (1.73)	37	2.62 (1.46)
Independence	34	1.80 (1.69)	30	2.81 (1.51)
General cognitive ability (intelligence)	35	1.70 (1.89)	17	3.28 (1.25)
Preliminary plans for the master's program (e.g., having an idea about the laboratory)	36.5	1.60 (1.35)	29	2.82 (1.41)
Maturity/responsibility/work habits	36.5	1.60 (1.71)	21	3.12 (1.50)
Integrity/fairness/honesty	38	1.50 (2.01)	23	3.06 (1.55)
Conscientiousness (trustworthiness and will to achieve)	39.5	1.40 (1.84)	25	3.02 (1.41)
Understanding of social relevance of research	39.5	1.40 (1.43)	33	2.75 (1.47)
Interpersonal skills (the ability to communicate or interact well with other people)	41.5	1.20 (1.32)	31	2.76 (1.40)
Experience in teamwork	41.5	1.20 (1.23)	32	2.75 (1.55)
Volunteering experience (amount and quality)	44	1.10 (0.88)	50	1.56 (1.11)
Openness to experience (imaginativeness, broad-mindedness, and artistic sensibility)	44	1.10 (1.45)	36	2.66 (1.48)
Emotional Stability	44	1.10 (1.60)	44	2.13 (1.50)
Time management skills / Timeliness	46.5	1.00 (1.33)	27	2.89 (1.61)
Leadership qualities	46.5	1.00 (1.70)	48	1.82 (1.33)
Travelling experience	49	0.80 (0.63)	51	1.56 (1.14)
Interests (sports, creative activities, etc.)	49	0.80 (.0.63)	47	1.99 (1.26)
Agreeableness (likability and friendliness)	49	0.80 (1.14)	46	2.03 (1.36)
Extraversion (activity and sociability)	51	0.70 (1.16)	49	1.79 (1.28)

"academic background" ($p = .001$), "grades" ($p = .009$), and "motivation factors" ($p < .001$). Also, the domain "motivation factors" was more important than the domain "cognitive ability" ($p = .041$).

Analysis of the GSNS reports from the admissions committee members showed that there was also a large effect of domain on average importance of criteria, $F(5, 43) = 8.77$, $p < .001$, $\eta^2 = .51$. Hochberg's post-hoc examination revealed that the "personality and personal competencies" domain was less important compared to three other domains: "academic background" ($p < .001$), "grades" ($p = .001$), and "motivation factors" ($p = .005$).

A two-way factorial ANOVA revealed that there was no significant difference between reports of the GSLS and the GSNS members of admissions committees, $F(5, 86) = 1.30$, $p = .271$, $\eta^2 = .07$.

Table 5 | *Means and Standard Deviations for Domains of Admissions Criteria as Reported by Admissions Committees Members and Perceived by Students at the Graduate School of Life Sciences (GSLS) and the Graduate School of Natural Sciences (GSNS).*

Domain of admissions criteria	GSLS				GSNS			
	Admissions committees members		**Applicants**		**Admissions committees members**		**Applicants**	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Academic background	2.89	0.89	3.21	0.48	3.03	0.97	3.14	0.49
Grades	3.05	0.42	2.90	0.53	3.05	0.58	2.68	0.55
Cognitive ability	2.32	0.46	3.27	0.61	2.15	0.38	3.12	0.68
Research background	2.66	0.56	3.24	0.22	2.38	0.76	2.90	0.26
Personality and personal competencies	1.75	0.39	2.83	0.55	1.27	0.52	2.52	0.62
Motivation factors	3.35	0.68	3.48	0.78	2.49	0.87	3.34	0.77

3.3.2 Research Question #2

Our results showed that at the GSLS, such selection methods as motivation letters, CVs, interviews, letters of recommendation, and personal acquaintance, if available, were all used extensively for assessment of admissions criteria[12]. For example, a motivation letter was used for assessment not only of criteria related to motivation itself, but also of criteria related to personality, cognitive ability, and previous academic background. Next, there were methods, namely, transcript and course content that were used in a more focused way. For example, course content was used predominantly for assessment of academic background, but not for assessment of personality and motivation. Finally, writing samples and English language tests were not frequently used: Only a few members of admissions committee (25%) indicated that they applied those methods.

At the GSNS, similarly to the GSLS, such selection methods as motivation letters, CVs, and letters of recommendation were used frequently and for measurement of diverse admissions criteria, while transcript and course content were also used extensively, but more focused (namely, for measurement of academic background). In contrast to the GSLS, interviews and inferences based on personal acquaintance were not frequently use. Similar to the GSLS, at the GSNS, an English language test and a writing sample were not often used for selective admissions decisions.

12 Frequencies are available upon request

3.3.3 Research Question #3

The importance of selective admissions criteria in the actual admissions decisions and transparency of selective admissions for the applicants were examined on two levels: on the level of specific selective admissions criteria and on the level of general admissions domains.

For the GSLS, the correlation between the importance of specific admissions criteria in decision-making as reported by admissions committee members and the perceived importance of the same criteria by the applicants, was of a moderate-to-large size, $r_{Pearson}$ (51) = .53, $p < .001$. Two-way analysis of variance on admissions domains revealed that there was significant interaction effect between the respondent group (admissions committee members versus applicants) and the type of domain (academic background, grades, cognitive ability, research background, personality and personal competencies, and motivation factors) on their importance, $F(5, 86) = 2.51$, $p = .036$, $\eta^2 = .13$. As seen in Fig. 2, the extent to which the GSLS applicants perceived criteria related to academic background, grades, research background, and motivation to be important was in coherence with reports from admissions committee members. However, the GSLS applicants overestimated the importance of criteria related to cognitive ability and personality.

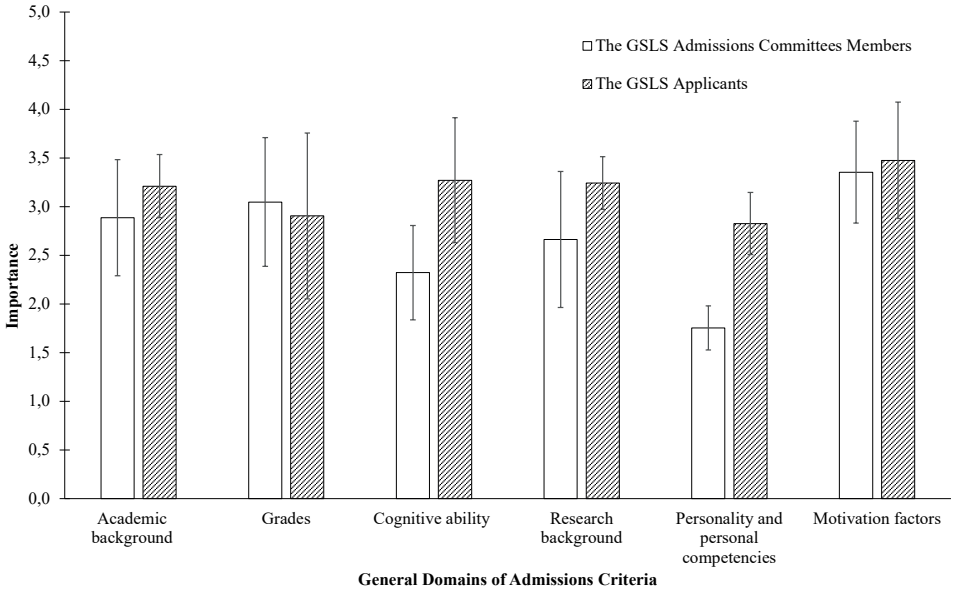


Figure 2 | *Importance of General Admissions Domains at the GSLS.*
Note. As reported by the GSLS admissions committee members (n = 16) and perceived by the GSLS applicants (n = 327). Means presented with their 95% CIs.

For the GSNS, similar results were obtained. Reports of admissions committee members on the importance of specific admissions criteria were correlated with the applicants' perceived importance to a high extent, $r_{Pearson}$ (51) = .63, $p < .001$. On the level

of admissions domains, two-way ANOVA revealed a significant interaction effect: The applicants perceived the importance of admissions criteria domains differently from what the admissions committee members reported, $F(5, 86) = 2.98$, $p = .016$, $\eta^2 = .15$. Figure 3 shows that four domains—academic background, grades, research background, and motivation—were perceived by the GSNS applicants to have similar importance in admissions decisions as was reported by the admissions committee members. Like their GSLS peers, the GSNS applicants overestimated the importance of criteria related to cognitive ability and personality.

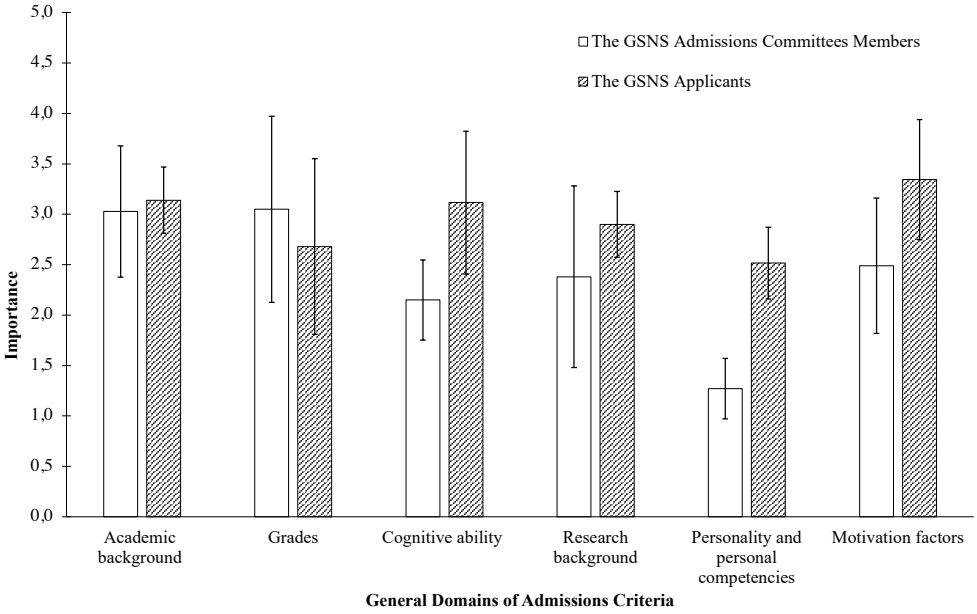


Figure 3 | *Importance of General Admissions Domains at the GSNS.*
Note. As reported by the GSNS admissions committee members ($n = 10$) and perceived by the GSNS applicants ($n = 153$). Means presented with their 95% CIs.

3.4 Discussion

In this study, we aimed to answer two central questions: How do admissions committees of the life and natural sciences research masters' programs select? Do applicants of two particular Life and Natural Graduate Schools know how the admissions committees select? We examined what selective admissions criteria are important in admissions decisions at the Life and Natural Sciences Graduate Schools. At both graduate schools, criteria related to academic background, research background, grades, cognitive ability, and motivation factors were reported by admissions committee members to be

important, but not the criteria related to personality and personal competencies. We also examined what kind of selection methods are used the most to assess selection criteria. We found that an unstructured motivation letter, a CV, and a transcript with grades are the most frequently used methods, however, other methods are used for measurement of various criteria as well. Finally, we investigated how accurate applicants, who have just undergone the admissions process, perceive the importance of admissions criteria in decisions regarding their application. We found that applicants' perceptions were accurate to a moderate degree. Below we discuss the evidence on the criteria and related selection methods per admissions domain.

3.4.1 Academic Background

Our findings suggest that a match between content of previous education and a master's program is by far one of the most important criteria in admissions decisions (rank #2 at the GSLS and rank #1 at the GSNS). This finding is consistent with other studies (Karazsia & McMurtry, 2012; Karazsia et al., 2013; Posselt, 2016; Walpole et al., 2002). Additionally, we find that criteria related to quality of undergraduate experience (a prior degree in a specific field, type of prior educational institution, and grades) are reported to be important in admissions decisions, which is also in line with existing literature (MasterMind Europe 2017; Kogan et al. 2015). To the best of our knowledge, there are no comprehensive studies on association between previous academic background (such as type of previous institution etc.) and graduate study success. Nevertheless, it seems plausible that prior academic background in a certain discipline (ranked #5 by the GSLS and #2 by the GSNS) provides an advantage for admissions to programs within a similar discipline. However, the high rankings of a criterion "type of prior educational institution" (a preference for either a Dutch research university/comparable international university or for university college or university of applied sciences (HBO)– rank #3 at both graduate schools) raises concerns, since giving a decision power to this criterion could place applicants from certain type of institutions in a dis/advantageous position. In order to account for differences in acquired level of knowledge and skills at different types of HEIs[13], one possible solution could be to provide an opportunity for the applicants to deliver the results of validated graduate standardized tests as evidence of their abilities, necessary for starting education at a graduate level, as there is evidence that standardized tests are valid predictors for graduate study success (e.g., Kuncel et al., 2001; Oh et al., 2008, and Chapter 5 of this thesis).

13 See Chapter 4 of this thesis, where we tested the predictive value of "type of prior education institution" toward graduate study success on the life sciences programs.

3.4.2 Grades

Both on a level of specific criteria and on a level of general domains, we find that criteria related to grades play an important role in admissions decisions in both graduate schools. For example, undergraduate GPA and grades for relevant courses are in the top-10 criteria at both Life and Natural Sciences Graduate Schools. On the level of more general domains, grades are shown to be of moderate importance in admissions decisions. Importance of grades in admissions decisions at both GSLS and GSNS can be regarded as a good practice, as prior findings indicate that undergraduate GPA—and last-year GPA in particular—are valid predictors of a number of future study success dimensions (Kuncel et al., 2001; Zimmermann, 2016).

3.4.3 Cognitive Ability

At both GSLS and GSNS, criteria related to cognitive ability are reported to be of a slight-to-moderate importance and are less important than motivation, academic background, research background, and grades. Inference about cognitive ability is done based on different selection methods (including motivation and recommendation letters, interviews, personal acquaintance). A variety of methods for inferring information about cognitive ability of an applicant raises concerns. Some of those selection methods were not developed for cognitive ability measurement, and they have serious issues with reliability and validity indicators (Patterson et al., 2016). The tests that are primarily developed for measurement aptitude and cognitive ability (e.g., Miller Analogies Test, Graduate Record Examinations, Graduate Management Admission Test) are not required at the examined programs.

3.4.4 Research Background

Not surprisingly, the current study also finds that criteria related to research experience play a role in admissions decisions. There are only a few studies in the life sciences field that investigated the predictive validity of amount of previous research experience and their findings are contradictory: While in one study a relation between amount of prior research experience and research productivity was found (Weiner, 2014), no relation was found in another study that tried to replicate the findings of the first study (Hall et al., 2017). While it is plausible to assume that previous research experience is associated with academic success at research-oriented masters' programs, more evidence is needed in order to use it as a selection criterion in admissions practice.

3.4.5 Motivation Factors

We found that motivation factors play an important role in admissions decisions: especially at the GSLS (six out of the top 10) and, to a lesser extent, at the GSNS (two out of the top 10). When we looked at the motivation factors as a general domain, we

saw that they were slightly more than moderately important at the GSLS and slightly-to-moderate important at the GSNS. Finally, criteria related to motivation factors were reported to be measured predominantly by unstructured motivation letters, CVs, interviews, and unstructured recommendation letters. While some motivational aspects were shown to have associations with study success (e.g., Robbins et al., 2004), operationalization of motivation in admissions raises serious concerns, as three indicated selection methods are self-reported measures.

Researchers strongly oppose using self-reported measures in student selection (Murphy et al., 2009; Niessen et al., 2017; Patterson et al., 2016; Wouters, 2016). It was demonstrated that self-reported measures induce a self-presentation effect, or, in other words, applicants tend to fake their reports in the selection context, which represent a high-stake situation (Niessen et al., 2017). Also, these measures have a small predictive value toward future study success and do not provide sufficient incremental validity in addition to prior grades and results of standardized admissions tests (Goho & Blackman, 2006; Hell et al., 2007; Murphy et al., 2009). As for the measure that is also extensively used for assessment of an applicant's motivation and is not self-reported—a recommendation letter—the scientific evidence exists for its practical value only in its standardized form (Kuncel et al., 2014).

3.4.6 Personality and Personal Competencies

Low importance of personality factors for admissions decisions is not in line with prior indications of the importance of personality and personal competencies in student selection (Conlon et al., 2012; Walpole et al., 2002). At least two possible explanations arise. The first is related to the design of our study since we used self-reports of admissions committee members and, therefore, the answers could be socially desirable. Other methods (e.g., observations) could shed light on the possible importance of an applicant's personality during the admissions process that may have been avoided in a self-report. Another possible explanation may be that for research-oriented masters' programs, the applicant's personality indeed does not play an important role in decision-making regarding their admissions. An interview study on selective admissions to undergraduate programs supports this explanation in the context of US and UK elite universities. The study found that admissions on academic record alone is typical for UK universities and is an exception at the US universities, where more attention is given to personal attributes of applicants (Mountford Zimdars, 2016). Also, the survey study of the MasterMind project showed that only one third of European programs consider personal skills in admissions decisions (MasterMind Europe, 2017). It is plausible to assume that the admissions practices in the Netherlands are more similar to other European countries' practices rather than to those of the US. In other words, according to the above-mentioned literature, the value that is assigned to personality and personal competencies in admissions decisions in European selective admissions is rather modest, and our results are in line with this literature.

A limited usage of personality traits and competencies in admissions decisions can be regarded as a good practice. One of the most stable findings is that a personality trait such as conscientiousness is associated with study success (Poropat, 2009; Schneider & Preckel, 2017; Trapmann et al., 2007), but the predictive validity of other personality traits—openness to experience, agreeableness, neuroticism, and extraversion—range from negligible to small, with an overall conclusion that they do not have a substantial impact on study success. It could be argued that results of assessment of conscientiousness can be used in admissions, however, since personality testing is usually self-reported and, therefore, susceptible for faking (Niessen et al., 2017), caution should be used when considering conscientiousness as selection criterion.

3.4.7 Are the Applied Admissions Criteria Transparent for Applicants?

We explored whether the applicants (who already knew the admissions decisions) accurately perceived the selection criteria. Our results show that although transparency is of a moderate-to-high extent on the level of specific criteria, on the level of general academic domains it is not the case. Applicants accurately perceive certain domains (academic background, grades, research background, and motivation factors), but not others (cognitive ability and personality are overestimated in terms of their importance in admissions decisions by applicants of both graduate schools).

These results suggest that some of the information that is presented to applicants (e.g., on the admissions webpages) provides a clear description of the criteria (e.g., minimum scores for English test, or a specific list of courses, which an applicant had to follow during previous education), but other information might be unclear (e.g., such criteria as "talent" or "high level of academic and professional ability"). Additionally, it could also be that applicants do not use enough information sources (such as a program's webinars, masters' open days, or even communication with a program coordinator via email) to inform themselves about the selection criteria. As one of the applicants commented: "I did not find much concrete accessible information on the selection criteria, but I did not look very intensely".

3.4.8 Strength, Limitations, Further Directions, and Conclusions

Among the strengths of this study is the availability of data from both admissions committee members and applicants. Additionally, data from two graduate schools with different scopes and curricula designs were available. The information from admissions committee members allowed us to describe the actual admissions practices, and information from the applicants allowed us to assess whether these practices are clear to those for whom they are designed. Next, we used a comprehensive list of specific admissions criteria which allowed us to conduct analysis both on a level of specific criteria and on a level of domains (when several similar criteria were considered together). With this approach, we were able to capture the specific trends in applicants' estimation of

admissions criteria importance. The fact that for both graduate schools we found either correct estimation or overestimation in the same admissions criteria domains speaks for a possible generalizability of our results to research masters' programs. Furthermore, to the best of our knowledge, this is the second study in the European context that explored the views on admissions practices from both admissions committee members and applicants (see MasterMind Europe, 2017, for the survey across Europe). This study, however, is the first that directly contrasted perceived importance against actual practices of selective admissions.

Despite these strengths, the consideration of some limitations should be made when interpreting the findings. First, a self-report by admissions committee members leaves a possibility that the reported importance of certain admissions criteria does not completely correspond with their real importance in the daily practice of selective admissions decision-making. Nonetheless, we assume that the staff members' reports closely resemble their practice. Secondly, in the third part of the study, dedicated to transparency of selective admissions criteria, most of the respondents were accepted applicants. Therefore, the results could have been affected by a selection bias: The accepted applicants might be those who had a chance to learn about the "unspoken" admissions rules in contrast to students who did not have a chance to learn such implicit information from their environment. Thirdly, the study is conducted within one HEI that could narrow down the generalizability. Yet, the study involves two large graduate schools (one of which is interfaculty) and multiple masters' programs in each school. Therefore, we expect the results to be relevant for (life and natural) science programs more generally.

In summary, this study contributes to our knowledge on what admissions committee members are looking for in future graduate students who apply for a research master's program. As a practical implication, the study suggests that the quality of admissions should be a subject of evaluation and constant improvement in terms of scientific evidence, fairness, and transparency. The results are relevant for applicants, researchers in higher education (in particular with a focus on admissions quality), and for professionals who are involved in the (daily) practice of selective admissions in HEIs.

3.5 References

Admissions to Higher Education Steering Group. (2004). *Fair admissions to higher education: Recommendations for good practice*. https://dera.ioe.ac.uk/5284/1/finalreport.pdf

Adnett, N., McCaig, C., Slack, K., & Bowers-Brown, T. (2011). Achieving transparency, consistency and fairness in English HE admissions: Progress since Schwartz*? Higher Education* Quarterly, 65, 12–33. https://doi.org/10.1111/j.1468-2273.2010.00468.x

Association of Universities in the Netherlands. (2018). *Aantal ingeschreven studenten and Internationale studenten [Number of enrolled students and international students]*. https://www.vsnu.nl/f_c_ingeschreven_studenten.html

Bagshaw, E. (2016, September 19). *Sydney university reveals real ATAR admissions scores*. The Sydney Morning Herald. Retrieved February 23, 2022, from https://www.smh.com.au/education/sydney-university-reveals-atar-admissions-scores-20160919-grjjie.html

Cleland, J. A., Dowell, J., McLachlan, J., Nicholson, S., and Patterson, F. (2012). *Identifying best practice in the selection of medical students: Literature review and interview survey*. https://www.gmc-uk.org/-/media/gmc-site-images/about/identifyingbestpracticeintheselectionofmedicalstudentspdf51119804.pdf?la=en&hash=D06B62AD514BE4C3454DEECA28A7B70FDA828715

Conlon, P., Hecker, K., & Sabatini, S. (2012). What should we be selecting for? A systematic approach for determining which personal characteristics to assess for during admissions. *BMC Medical Education, 12,* 1–10. https://doi.org/10.1186/1472-6920-12-105

Craig, P. L., Gordon, J. J., Clark, R. M., & Langendyk, V. (2004). Prior academic background and student performance in assessment in a graduate entry programme. *Medical Education, 38,* 1164–1168. https://doi. org/10.1111/j.1365-2929.2004.02043.x

Goho, J., & Blackman, A. (2006). The effectiveness of academic admission interviews: An exploratory meta-analysis. *Medical Teacher, 28,* 335–340. https://doi.org/10.1080/01421590600603418

Graduate School of Life Sciences (n.d.). *Mission of the GSLS*. Retrieved February 23, 2022, from https://www.uu.nl/en/education/graduate-school-of-life-sciences/about-us/mission-of-the-gsls-0

Hall, J. D., O'Connell, A. B., & Cook, J. G. (2017). Predictors of student productivity in biomedical graduate school applications. *PLoS One, 12*(1), Article e0169121. https://doi.org/10.1371/journal.pone.0169121.

Hell, B., Trapmann, S., Weigand, S., and Schuler, H. (2007). Die Validität von Auswahlgesprächen im Rahmen der Hochschulzulassung -- eine Metaanalyse [The validity of admission interviews -- a meta-analysis]. *Psychologische Rundschau, 58*(2), 93-102. https://doi. org/10.1026/0033-3042.58.2.93

Karazsia, B. T., & McMurtry, C. M. (2012). Graduate admissions in pediatric psychology: The importance of undergraduate training. *Journal of Pediatric Psychology, 37*, 127–135. https://doi.org/10.1093/jpepsy/jsr067

Karazsia, B. T., Stavnezer, A. J., & Reeves, J. W. (2013). Graduate admissions in clinical neuropsychology: The importance of undergraduate training. *Archives of Clinical Neuropsychology, 28*, 711–720. https://doi.org/10.1093/arclin/act056

Kogan, L. R., Stewart, S. M., Schoenfeld-Tacher, R., and Hellyer, P. W. (2015). Perceptions of veterinary admissions committee members of undergraduate credits earned from community colleges or online compared to traditional 4-year institutions. *Open Veterinary Journal, 5*, 71–84. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4629565/

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127*, 162–181. https://doi.org/10.1037/0033-2909.127.1.162.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86*, 148–161. https://doi.org/10.1037/0022-3514.86.1.148.

Kuncel, N. R., Kochevar, R. J., & Ones, D. S. (2014). A meta-analysis of letters of recommendation in college and graduate admissions: Reasons for hope. *International Journal of Selection and Assessment, 22*, 101–107. https://doi.org/10.1111/ijsa.12060.

la Noue, G. R. (2003). The new imperative for admissions transparency. *Academic Questions, 16*(3), 23–29. https://doi.org/10.1007/s12129-003-1076-9.

MasterMind Europe. (2017). *Admissions to English-taught programs (ETPs) at Master's level in Europe—Procedures, regulations, success rates and challenges for diverse applicants*. Amsterdam: ACA, StudyPortals, and Vrije Universiteit Amsterdam. http://mastermindeurope.eu/wp-content/uploads/2017/01/Report-2-Admissions-to-ETPs.pdf

Mountford Zimdars, A. (2016). *Meritocracy and the university: Selective admission in England and the United States*. London; New York: Bloomsbury Academic.

Murphy, S. C., Klieger, D. M., Borneman, M. J., & Kuncel, N. R. (2009). The predictive power of personal statements in admissions: A meta-analysis and cautionary tale. College & University, 84(4), 83–88. https://www.proquest.com/docview/225605798

Nauta, M. M. (2000). Assessing the accuracy of psychology undergraduates' perceptions of graduate admission criteria. *Teaching of Psychology, 27*(4), 277–280.

Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2017). Measuring non-cognitive predictors in high-stakes contexts: The effect of self-presentation on self-report instruments used in admission to higher education. *Personality and Individual Differences, 106*, 183–189. https://doi.org/10.1016/j.paid.2016.11.014

Nuffic (2015). *Education system in the Netherlands*. https://www.nuffic.nl/sites/default/files/2020-08/education-system-the-netherlands%20%281%29.pdf

Oh, I., Schmidt, F. L., & Shaffer, J. A. (2008). Admission test (GMAT) is even more valid than we thought: A new development in Meta- analysis and its implications for the validity of the GMAT. *Academy of Management Learning and Education, 7*, 563–570. https://doi.org/10.5465/amle.2008.35882196.

Pacheco, W. I., Noel, R. J., Porter, J. T., & Appleyard, C. B. (2015). Beyond the GRE: Using a composite score to predict the success of Puerto Rican students in a biomedical PhD program. *CBE—Life Sciences Education*, *14*(2), Article14:ar13. https://doi.org/10.1187/cbe.14-11-0216

Patterson, F., & Ferguson, E. (2010). Selection for medical education and training. In T. Swanwick (Ed.), *Understanding medical education: Evidence, theory and practice* (pp. 352–365). https://doi.org/10.1002/9781444320282.ch24

Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., & Cleland, J. (2016). How effective are selection methods in medical education? A systematic review. *Medical Education*, *50*(1), 36–60. https://doi.org/10.1111/medu.12817

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*(2), 322–338. https://doi.org/10.1037/a0014996

Posselt, J. R. (2016). *Inside graduate admissions: Merit, diversity, and faculty gatekeeping*. Harvard University Press.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A Meta-analysis. *Psychological Bulletin, 130*, 261–288. https://doi.org/10.1037/0033-2909.130.2.261.

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin, 143*, 565–600. https://doi.org/10.1037/bul0000098.

Kira Talent (2018). *Breaking down bias in admissions: The how-to guide to preventing admissions bias at your school*. http://start.kiratalent.com/breaking-down-admissions-bias/

The Dutch National Students Association (Interstedelijk Studenten Overleg). (2014). *Meer transparantie bij decentrale selectie. Het belang en een framework [More transparency in decentralised selection. Importance and a framework]*. https://issuu.com/interstedelijkstudentenoverleg/docs/1415_meer_transparantie_bij_decentr

Tillema, H., Leenknecht, M., & Segers, M. (2011). Assessing assessment quality: Criteria for quality assurance in design of (peer) assessment for learning - a review of research studies. *Studies in Educational Evaluation, 37*, 25–34. https://doi.org/10.1016/j.stueduc.2011.03.004.

Times Higher Education (2018, October 1). Best universities in the Netherlands 2019. https://www.timeshighereducation.com/student/best-universities/best-universities-netherlands

Trapmann, S., Hell, B., Hirn, J.-O. W., & Schuler, H. (2007). Meta-analysis of the relationship between the big five and academic success at university. *Zeitschrift Für Psychologie, 215*, 132–151. https://doi.org/10.1027/0044-3409.215.2.132

Utrecht University. (2017). *Utrecht University Education Guideline.* https://www.uu.nl/sites/default/files/education_guideline_2017.pdf

van Ooijen-van der Linden, L., van der Smagt, M. J., Woertman, L., & te Pas, S. F. (2017). Signal detection theory as a tool for successful student selection. *Assessment and Evaluation in Higher Education, 42*, 1194–1207. https://doi.org/10.1080/02602938.2016.1241860

von Stumm, S., Hell, B., & Chamorro-Premuzic, T. (2011). The hungry mind: Intellectual curiosity is the third pillar of academic performance. *Perspectives on Psychological Science, 6*, 574–588. https://doi.org/10.1177/1745691611421204

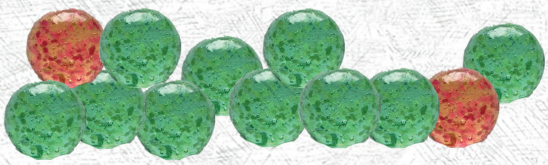
Walpole, M., Burton, N. W., Kanyi, K., & Jackenthal, A. (2002). *Selecting successful graduate students: in-depth interviews with GRE ® users.* (Report No. GREB-99-11R, RR-02-08). Princeton: Educational Testing Service. https://doi.org/10.1002/j.2333-8504.2002.tb01875.x

Weiner, O. D. (2014). How should we be selecting our graduate students? *Molecular Biology of the Cell, 25*, 429–430. https://doi.org/10.1091/mbc.E13-11-0646

Wouters, A. (2016). *Effects of medical school selection on the motivation of the student population and applicant pool* [Doctoral dissertation, VU Amsterdam]. http://dare.ubvu.vu.nl/handle/1871/55083

Zimmermann, J. (2016). *Information Processing for Effective and Stable Admission* [Doctoral dissertation, ETH Zürich]. https://doi.org/10.3929/ethz-a-010664128

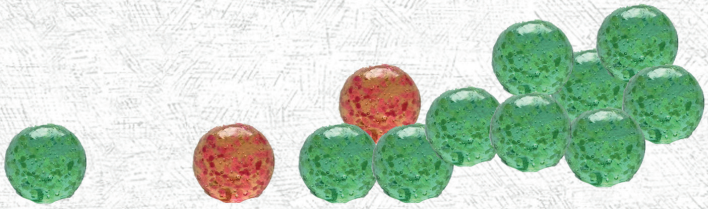
4



Acknowledgements

The authors would like to thank educational data specialist Guus Dikker. His professional advice on data processing was very valuable.

Once the Best Student Always the Best Student? Predicting Graduate Study Success Using Undergraduate Academic Indicators: Evidence from Research Masters' Programs in the Netherlands



Author note

An adapted version of this chapter is published as "Kurysheva A., Koning N., Fox C. M., van Rijen H. V. M., & Dilaver G. (2022). Once the Best Student Always the Best Student? Predicting Graduate Study Success Using Undergraduate Academic Indicators: Evidence from Research Masters' Programs in the Netherlands. *International Journal of Selection and Assessment*. 1–17. https://doi.org/10.1111/ijsa.12397"

A. Kurysheva, N. Koning, H. V. M. van Rijen, and G. Dilaver conceptualized the overarching research aims. A. Kurysheva developed the methodology, conducted a portion of the cleaning of the data, conducted formal analysis, and prepared the original draft. N. Koning developed a procedure for grade conversions, conducted a portion of the cleaning of the data, and created a query for downloading the data from the administrative system. C. Fox provided substantive editing for the entire text, contributed to the improvement of its structure, content, and flow. All authors provided critical feedback on the manuscript.

Abstract

In the face of increasing and diversifying graduate application numbers, evidence-based selective admissions have become a pressing issue. By conducting multilevel regression analyses on institutional admissions data from a Dutch university, this study aims to determine the predictive value of undergraduate academic indicators for graduate study success on research masters' programs in the life sciences. The results imply that in addition to undergraduate grade point average (UGPA), undergraduate thesis grade is a valid predictor of graduate grade point average (GGPA). To a small extent, the examined undergraduate academic indicators also predict graduate degree completion and time to degree. The results from this study can be used by admissions committees for evaluating and improving their current practices of graduate selective admissions.

Keywords: admissions, student selection, graduate education, master's degree, study success, GPA, thesis

4.1 Introduction

The goal of university admissions committees is to create a selective admissions process that meets societal expectations of objectiveness, fairness, and transparency. Over the last two decades, countries with widespread instruction in English have seen a steady increase in demand for graduate education (Association of Universities in the Netherlands, 2021; Statista, 2020; the Higher Education Statistics Agency, 2020). This demand has challenged admissions committees for several reasons. First, many universities now face a disparity in the number of graduate school places versus the number of applicants. This has created a situation where some students, whilst eligible, are rejected. Admissions committees, therefore, must be able to justify their selection decisions. Second, because of the growing number of internationally mobile students (Organisation for Economic Co-operation and Development [OECD], 2022), application files have become more diverse. Admissions committees are now faced with the challenge of comparing foreign applications (from different education systems with different evaluations processes) against national students. Third, despite efforts to increase access to higher education for underrepresented groups such as first-generation students, students with disabilities, and students with migration backgrounds (Torgerson et al., 2014; Younger et al., 2018; Torotcoi et al., 2020), these groups still have less chance of accessing Higher Education Institutions (HEIs; Salmi & Bassett, 2014). For these reasons, university admissions committees need valid selection methods.

HEIs implement an array of selection methods for making admissions decisions. Some use information on applicants' prior education and performance. We refer to this as *undergraduate academic indicators*. Some other selection methods are based on additional assessments. These are done either by third parties (e.g., a referee, a testing center) or by applicants themselves (e.g., personal statements). In this article, we focus on undergraduate academic indicators because they are typically available in any admissions institutional data. Using these data in our research enlarges the possible generalizability of study findings across different HEIs.

The importance of undergraduate academic indicators in making decisions about admissions varies. According to surveys across European universities, the most widely used indicators for selection to masters' programs are *undergraduate grade point average (UGPA), undergraduate research/work experience, type of prior HEI*, and *prior field of study* (Kurysheva et al., 2019; MasterMind Europe, 2017). These indicators have also been shown to be important in admissions decisions in the United States (US; Kogan et al., 2015; Walpole et al., 2002).

Existing evidence suggests that among the four undergraduate academic indicators mentioned above, only UGPA is consistently found as a valid predictor of graduate study success (Kuncel et al., 2007; Schneider & Preckel, 2017). However, UGPA predicts different

graduate study success dimensions to a different extent. The evidence also shows that higher UGPA is related to higher graduate grade point average (GGPA; Burton & Wang, 2005; Fu, 2012; Howell et al., 2014; Moneta-Koehler et al., 2017; Zimmermann et al., 2015), but when it comes to other dimensions of graduate study success, the evidence is mixed. For instance, some studies also found a positive relationship between UGPA and graduate degree completion (e.g., Moneta-Koehler et al., 2017; Schwager et al., 2015; Wollast et al., 2018), while others did not (Cox, Hughes, et al., 2009; Dore, 2017). The same applies for graduate time to degree: Some studies found a negative relationship between UGPA and graduate time to degree (Howell et al., 2014) and others did not (Dabney, 2012; Moneta-Koehler et al., 2017). We, therefore, expect that UGPA should positively relate to GGPA, but with less certainty to other dimensions of graduate study success, and even if it predicts them, we expect that the extent of this prediction to be small (Hypothesis 1).

For undergraduate research/work experience, the meta-analytical evidence shows that it is largely unrelated to academic performance in the graduate phase but with less statistical certainty to degree attainment and publication performance (Miller et al., 2021). This was an unanticipated finding, considering the theoretical underpinnings of the critical importance of undergraduate research for graduate school performance in research skills (Gilmore et al., 2015). The findings in the meta-analysis of Miller et al. (2021) could be explained by the fact that more than a half of the included studies used the generic dichotomous operationalization of undergraduate research experience ("present or absent"). When undergraduate research experience is operationalized differently (e.g., as "duration of research experience in months" or "whether or not the student wrote a thesis during their bachelor's program"), some studies find the relationship between research experience and certain dimensions of graduate study success (see Gilmore et al., 2015, Cox et al., 2009, Weiner, 2014), while others do not (Hall et al., 2017). There have been no studies conducted on the relationship between undergraduate thesis grade and GGPA. However, out of all undergraduate study activities, undergraduate thesis is the most pertinent and usually also the most recent indicator in relation to research-oriented graduate education. Hence, we hypothesize that undergraduate thesis grade should positively relate to graduate study success on research-intense masters' programs (Hypothesis 2).

The other commonly used undergraduate academic indicators (type of prior HEI and prior field of study) have not been studied in meta-analyses and only a handful of primary studies exist that address their predictive validity. A study on a German sample of business administration and economics graduate students found that a student's former HEI had a substantial effect on graduate study success, however, the type of HEI had a weak effect on graduate study success (categorized as university, college [Fachhochschule], academy [Berufakademie], and school abroad; Chadi & de Pinto, 2017). A study on an Australian sample of medical students showed a negligible

difference in students' performance on the assignments based on their prior field of study (Craig et al., 2004). There is obviously not enough research on these two academic indicators to come to an established hypothesis. This study will address this gap in research; however, we do not expect type of prior HEI and prior field of study to be strong determinants of graduate study success (Hypotheses 3 and 4, respectively).

To enrich our knowledge on predictive validity of commonly used undergraduate academic indicators overall, it is important to study the type of prior HEI and field of study of incoming graduate students. It is also important to search for new operationalizations of those undergraduate academic indicators that are theoretically sound and extensively used by the admission committees but have not been validated (i.e., new operationalizations for research experience). Cognizant of the fact that there are other multiple considerations when it comes to determining whether an undergraduate academic indicator is suitable for use in student selection (see Patterson et al., 2016, 2018; Posselt, 2016), this study focuses on predictive validity.

4.1.1 Current Study

In this study, we examine to what extent the four most widely used undergraduate academic indicators predict graduate study success of research masters' programs in the life sciences. Half of the studies reviewed were conducted on taught masters' programs – programs with course-based curricula (AHZ Associates, 2021; The Quality Assurance Agency for Higher Education, 2020). The other half were conducted on research graduate programs (most of them were on doctoral level) which aim to develop advanced research skills (AHZ Associates, 2021; Snijder, 2016; The Quality Assurance Agency for Higher Education, 2020). We chose to focus on the research masters' programs (in contrast to taught masters' programs) in the life sciences because of the intensive study loads in research laboratories which require extensive and often long-lasting immersion in research practice.

The examined undergraduate academic indicators were: UGPA, undergraduate thesis grade as an operationalization of research experience, type of prior HEI, and prior field of study. We operationalized research experience through the undergraduate thesis grade due to its relative objectivity: It represents the quantitative assessment by experts in the field and this assessment often follows a certain rubric or at least requires certain extent of justification. An undergraduate thesis is a common part of undergraduate curriculum. Therefore, it makes it easier to place undergraduate thesis grade on a common metric, applicable for a substantial number of applicants, in contrast to other operationalizations of research experience (Miller et al., 2021). We operationalized graduate study success through three dimensions: (1) *graduate degree completion*, (2) *GGPA*, and (3) *graduate time to degree* (i.e., time taken to complete a master's degree). In addition to examining the direct relationships between four undergraduate academic indicators and dimensions of graduate study success, we also controlled for gender

and age (as the last step of our analysis) to see if the statistical isolation of their effects from the effects of the undergraduate academic indicators would change the results. The analysis with covariates included did not change the results substantially and is available by request.

4.2 Method

This quantitative study examines the relationship between student undergraduate academic indicators and their graduate study success. The goal of this study is to help provide guidance for graduate school admissions committees regarding which undergraduate academic indicators should be considered in student selection. To better understand the generalizability of this study and to set it within the context of other graduate programs, a national and institutional context is provided below.

4.2.1 National Context

This study has been conducted in a large research university in the Netherlands. The Dutch higher education system is comprised of fourteen public research universities that grant academic degrees up to the PhD level (including some university colleges which offer selective international liberal arts and sciences bachelors' programs), 37 universities of applied sciences (which grant professional degrees up to master's level), and a few small specialized private institutions (van der Wende, 2020). At research universities, research-intensive education aims to advance understanding of the phenomena studied within academic disciplines, to facilitate application of scientific knowledge, and to generate new knowledge. Universities of applied sciences offer higher professional education–theoretical and practical training related to professions that necessitate a higher vocational qualification (Eurydice, 2020).

In this article, we focus on masters' programs at research universities. For comparison, the Netherlands has adapted the Framework for Qualifications of the European Higher Education Area (QF-EHEA) which consists of three cycles (Bachelor's/Master's/PhD). It was introduced with the Bologna Process in 2002 (Lub et al., 2003; Witte et al., 2008) and covers levels 6-8 in the European Qualifications Framework. This means that the master's phase in the Netherlands is comparable to a master's phase in 48 countries within the EHEA (EHEA, n.d.). This three cycles framework is also compatible to both the US and Canada with only subtle differences with the UK which also offers an MPhil option that sits between a Master's and PhD.

It is possible to enter a Dutch master's program in a research university with an undergraduate degree either from a Dutch research university, university college, university of applied sciences, or the equivalent from a foreign HEI. Dutch research universities offer not only taught, but also research masters' programs. Research masters'

programs differentiate themselves from taught masters' with an emphasis on research, duration (two years and 120 EC instead of one year and 60 EC), and selective admissions of students (Snijder, 2016). They aim to prepare students for research-related positions both inside and outside academia (NVAO, 2016). The curriculum of these programs is specifically focused on obtaining and practicing research competencies and skills. For example, internships at research laboratories typically constitute components of research masters' programs in the life sciences.

4.2.2 Institutional context

We used data from an interdisciplinary graduate school of a major Dutch research university with thirteen RM programs in the life sciences. At this graduate school, the demand for study placement increases annually. The major research project of nine months represents the main component of the graduate curriculum. The remaining part of the curriculum consists of a minor research project, different mandatory and optional courses, and a writing assignment. The weighted grade for these components constitutes GGPA. The research projects are usually conducted in the university's laboratories. Students are exposed to a variety of research processes and are expected to conduct their own research that involves multiple stages, starting from research design and data collection to writing a research report.

4.2.3 Participants

No recruitment was needed because we used the institutional data (i.e., data from the university administrative system). This data usage was approved by the Netherlands Association for Medical Education Ethical Review Board (dossier number: 2019.8.2). The data came from six cohorts of 1,792 masters' students. Out of these students, 1,570 (88%) completed their masters' studies and 222 (12%) dropped out at some point during their masters' programs.

Out of the sample of 1,792 students, which is labelled Sample 1, three additional analytical sub-samples were derived (Sample 2, 3, and 4). Sample 1 and Sample 2 were used to predict the binary variable graduate degree completion. Sample 1 consisted of students who came from four different types of undergraduate HEIs ($N_{completed\_and\_droppedout\_from\_different\_HEI}$=1792). Sample 2 consisted of students who studied their masters' at the same university as their bachelors'; therefore, their undergraduate thesis grade was available[14] ($N_{completed\_and\_droppedout\_\&the\_same\_HEI}$=1249). Sample 3 and Sample 4 were used to predict two metric variables *(*GGPA and graduate time to degree*)*. These study success dimensions were only available for students who completed their studies.

14 Ideally, we would have wanted to use undergraduate thesis grade as a predictor in all our analyses. Unfortunately, these grades were not registered in the administrative system for students who had come from different universities. They were registered only for students who studied their bachelor's program at the same university as their master's program. For this reason, we had to analyze four samples instead of two.

Sample 3 consisted of students who came from four different types of undergraduate HEIs ($N_{completed\_from\_different\_HEI}$ =1570). Sample 4 consisted of students who studied their masters' at the same university as their bachelors'; therefore their undergraduate thesis grade was available ($N_{completed\_and\_droppedout\_\&the\_same\_HEI}$=1112). Information on sample sizes and characteristics is presented in Table 1.

Table 1 | *Samples' Demographical and Educational Characteristics.*

Characteristics	**Sample 1. Graduates and drop-outs from different prior HEIs (*N* = 1792)**	**Sample 2. Graduates and drop-outs who studied their masters' at the same university as their bachelors' (*N* = 1249)**	**Sample 3. Graduates from different prior HEIs (*N* = 1570)**	**Sample 4. Graduates who studied their masters' at the same university as their bachelors' (*N* = 1112)**
Gender: males (*n*)	741	521	627	456
Age range in years	17–49	17–49	18–38	19–38
M_{age}	22.5	22.3	22.4	22.2
SD_{age}	2.1	2.0	1.9	1.8
Mdn_{age}	22.0	22.0	22.0	22.0
Citizenship (%)				
The Netherlands	92	98	91	98
Other EU	8	1	8	1
Outside of EU	< 0.1	< 1	< 1	<1
Type of prior HEI (%)				
Dutch research university	84	100	84	100
Dutch university college	4		4	
Dutch university of applied sciences	7		7	
International HEI	5		5	
Prior field of study (%)				
Biology	34	41	34	41
Biotechnology	2		2	
Biology and medical laboratory	5		5	
Biomedical sciences	31	36	32	36
Chemistry	5	6	5	6
Liberal arts and sciences	6	1	5	
Medicine	1	1	1	1
Pharmaceutics	5	6	5	6
Psychology	5	5	5	6
Other	6	4	6	4
Missingness (values; %)	1.7	4.0	1.5	3.6

4.2.4 Measures

4.2.4.1 Independent Variables

Percentile Ranks of Undergraduate Grade Point Average (UGPA). UGPA refers to an average grade for all curriculum components of an undergraduate program, weighted according to the number of credits for each component. The UGPA of each student was transferred to the percentile ranks due to different grading systems that are applied at different Dutch and international education systems. Percentile ranks allowed us to place all student grades from different grading systems on one scale. The adequacy of usage of percentile ranks was double-checked via a stability check of results, using UGPA on a US scale (from 0 till 4) instead of percentile ranks.

The percentile ranks placed each student in a relative position to others from their own country. We used the data only from the largest groups ($n >= 20$), so that percentile ranks could be derived. Among the Dutch students, the percentile ranks were given within three groups: students from Dutch university colleges (UGPAs on a scale from 1 to 4), Dutch research universities (UGPAs on a scale from 1 to 10), and universities of applied sciences (UGPAs on a scale from 1 to 10). The largest international student groups, who were greater than or equal to twenty in size, came from the European Union (EU). Namely, the international student groups included British (UGPAs on a scale from 0 to 100), Greek (UGPAs on a scale from 1 to 10), Italian (UGPA on a scale from 0 to 30), and Spanish (UGPAs on a scale from 1 to 10). Other EU student groups and student groups outside of the EU were left out of the analysis due to insufficient numbers per group.

Undergraduate Thesis Grade. This variable (on the Dutch grading scale from 1 to 10) represents a grade for an undergraduate thesis or research project.

Prior Field of Study. This variable is nominal and indicates the major which a student specialized in during their undergraduate studies (see Table 1 for the list of fields of studies and respective frequencies).

Type of Prior Higher Education Institution. This variable is also nominal and represents types of HEIs where students completed their undergraduate programs. In our data, four types of programs were distinguished: Dutch research universities, Dutch university colleges, Dutch university of applied sciences, and international HEIs (see Table 1 for frequencies of each). The international HEIs were considered as one category. This is because in this specific sample only applications with the type of prior HEI—comparable to a Dutch research university—are usually processed further by the admissions committees. International students with an undergraduate degree from a HEI that is on the level of the Dutch universities of applied sciences are rarely ever admitted. Likewise, it is not common to admit students with an undergraduate degree from international colleges with liberal arts and sciences degrees (which would be an analogue to the Dutch university colleges). Therefore, in terms of the level of their prior type of HEI, the group of international students can be considered comparable

to the group of students from Dutch research universities. It was then decided to keep students from international HEIs as one group, which is in line with other studies in the field (e.g., Chadi & de Pinto, 2018).

4.2.4.2 Dependent Variables

Graduate Degree Completion. Graduate degree completion is a binary variable wherein the category "master's degree attained" (coded as 1) was defined as obtaining a master's degree within four years after the start of the master's program and a category "master's degree was not attained" was defined as an actual stoppage with the master's program (coded as 0).

Graduate Grade Point Average (GGPA). GGPA (on the Dutch grading scale from 1 to 10) represents an average grade for all curriculum components of a research master's program weighted according to their credit value.

Graduate Time to Degree. Graduate time to degree is measured as actual duration in months of the master's studies for each student. The expected duration on the research masters' programs at this graduate school is 24 months. However, students are allowed to graduate earlier or later, and it is common in this graduate school to graduate a few months later than the nominal duration of 24 months. Graduate time to degree in our student sample ranged from 19 to 84 months with a median of 28 months.

4.2.4.3 Research Model and Data Analysis Approach

Figure 1 presents our research model. It shows the examined relationships as well as intercorrelations between the predictors. Though the variables of interest are on an individual level, the data have a multilevel structure (students nested in 68 study groups which in turn are nested in thirteen programs). To account for the dependency of students within groups and programs, the hierarchical linear modelling was applied. We ran the analyses on four different dimensions of graduate success separately and not on one multivariate outcome because such a multivariate outcome would make the interpretation of findings barely explainable and, therefore, useless for admission practitioners. Analysis was conducted in HLM 8. The stability check of obtained results was also conducted using UGPAs converted to GPAs on the US scale instead of percentile ranks. Since the percentage of missingness was low (in all four samples less than 5% of data was missing), we handled the missingness using the Expectation-Maximization (EM) method.

Below, we describe the results for the incremental validity of type of prior HEI, undergraduate thesis grade, and field of study above and beyond UGPA for each of the graduate study success dimension. Table 3.1, Table 4.1, and Table 5.1 are based on analyses of Sample 1 and Sample 3 which included students from all types of HEIs. Table 3.2, Table 4.2, and Table 5.2 are based on analyses of Sample 2 and Sample 4 which included students who did their masters' at the same university as their bachelors'; therefore, their undergraduate thesis grade was available.

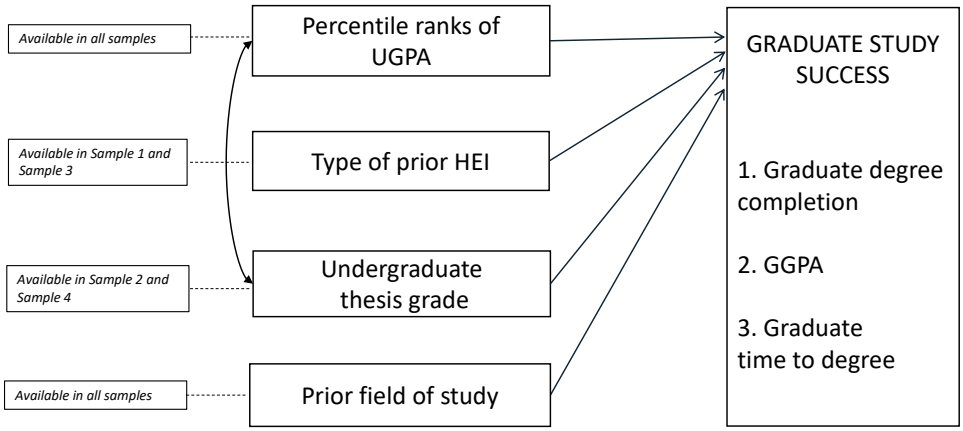


Figure 1 | *The Model with Undergraduate Academic Indicators – Predictors of Graduate Study Success.*

4.3 Results

Table 2 shows the Pearson correlations between the study variables. Both percentile rank of UGPA and undergraduate thesis grade are significantly related to the three dimensions of graduate study success: positively to degree completion and GGPA and negatively to graduate time to degree.

Table 2 | *Intercorrelations between Study Variables.*

Variable	*n*	*M*	*SD*	1	2	3	4
Sample 1							
1. Percentile rank of UGPA	1689	49.82	28.71	1			
2. Degree completion	1792	0.88	0.33	.053*	1		
Sample 2							
1. Percentile rank of UGPA	1186	49.54	28.83	1			
2. Undergraduate thesis grade	1011	7.76	0.72	.52***	1		
3. Degree completion	1249	0.89	0.31	.06	.09**	1	
Sample 3							
1. Percentile rank of UGPA	1186	49.54	28.83	1			
2. GGPA	1206	7.83	0.58	.53***	1		
3. Graduate time to degree	1090	30.72	7.93	-.13***	-.27***	1	
Sample 4							
1. Percentile rank of UGPA	1059	50.09	28.84	1			
2. Undergraduate thesis grade	902	7.78	0.71	.52***	1		
3. GGPA	1099	7.87	0.55	.58***	.47***	1	
4. Graduate time to degree	1099	30.72	7.93	-.13***	-.15***	-.27***	1

Note. Type of prior HEI and field of study are multinominal variables; therefore, they could not be included into the correlational table.
*$p < .05$. **$p < .01$. *** $p < .001$.

4.3.1 Graduate Degree Completion as a Dependent Variable

Table 3.1 shows that the result from uncorrected correlations—UGPA as a significant predictor of degree attainment—holds even after accounting for the dependency of students within groups and programs by applying hierarchical modelling (Model 1). Further models provide the results for the incremental validity of type of HEI and field of study above and beyond UGPA. Model 2 shows that students from Dutch research universities and from foreign HEIs have higher odds of completing a graduate program compared to students from Dutch universities of applied sciences and students from Dutch university colleges. Model 3 shows that undergraduate field of study of students is not related to their odds of completing a graduate degree. We also note that with adding each predictor (except field of study dummies), the Akaike Information Criterion (AIC)—an indicator of relative quality of statistical models—improves but rather to a small extent.

Table 3.1 | *Hierarchical Regression Results for Graduate Degree Completion. Sample 1.*

Variable	Model 0	Model 1	Model 2	Model 3
		Fixed effects		
Intercept	7.35*** [6.00, 9.01]	7.43*** [6.11, 9.05]	4.39*** [2.64, 7.32]	4.37** [1.50, 12.76]
Percentile rank of UGPA		1.01* [1.00, 1.01]	1.01* [1.00, 1.01]	1.01** [1.00, 1.01]
Dummies (Type of prior HEI)[a]				
Dutch research university			1.81* [1.13, 2.92]	1.60 [0.62, 4.13]
Dutch university college			0.82 [0.40, 1.67]	0.98 [0.19, 5.08]
Foreign HEI			3.91** [1.42, 10.78]	4.08* [1.17, 14.25]
Dummies (Prior field of study)[b]				
Biotechnology				0.78 [0.21, 2.80]
Biology and medical laboratory				0.99 [0.33, 2.97]
Biomedical sciences				1.33 [0.90, 1.96]
Chemistry				1.16 [0.57, 2.34]
Liberal arts and sciences				0.84 [0.24, 2.94]
Medicine				0.55 [0.17, 1.72]
Pharmaceutics				1.97 [0.82, 4.75]
Psychology				1.04 [0.48, 2.26]
Other				0.92 [0.50, 1.71]
		Random effects		
Variance components				
Level 1	3.29	3.29	3.29	3.29
Level 2	0.07	0.08	0.08	0.05
Level 3	0.01	< 0.01	< 0.01	< 0.01
		Goodness of fit		
Deviance	4633.71	4627.82	4612.15	4605.87
Number of estimated parameters	3	4	7	16
Model comparison test		$\chi^2(1) = 5.89$*	$\chi^2(3) = 15.67$**	$\chi^2(9) = 6.28$
AIC	4639.71	4635.82	4626.15	4637.87

Note. The reported estimates of predictors are odds ratios. Confidence intervals are in square brackets.
[a]The reference category: "Dutch university of applied sciences". [b]The reference category: "Biology". Rerunning analysis to test other dummies of types of prior HEI in Model 2 delivers also other significant differences, namely for a dummy variable "Dutch university college versus Dutch research university [ref]", Exp(b) = 0.45**, CI = [0.26, 0.80] and for "Foreign HEI versus Dutch university college [ref]", Exp(b) = 4.77**, CI = [1.65, 13.76]. In Model 3, there is no significant effect of a dummy variable "Dutch university college versus Dutch research

university [ref]" in a model with the dummies for field of study included, Exp(b) = 0.61, CI = [0.16, 2.35]. Likewise, there is no significant effect of the dummy variable "Foreign HEI versus Dutch university college [ref]", when dummies for field of study are added: Exp(b) = 4.14, CI = [0.79, 21.89].

*p < .05. **p < .01. ***p < .001.

Table 3.2 shows that once thesis grade is added to the model with UGPA, the model fit increases significantly, though the estimate for undergraduate thesis grade does not reach the chosen alpha level of 0.05. The improvement in AIC is small.

Table 3.2 | *Hierarchical Regression Results for Graduate Degree Completion. Sample 2.*

Variable	Model 0	Model 1	Model 2	Model 3
		Fixed effects		
Intercept	8.95*** [6.79, 11.80]	9.17*** [6.92, 12.14]	9.28*** [7.00, 12.32]	8.48*** [5.72, 12.56]
Percentile rank of prior average grade		2.01* [1.00, 1.02]	1.00 [1.00, 1.01]	1.00 [0.97, 1.01]
Thesis grade			1.38 [1.00, 1.91]	1.37 [0.99, 1.90]
Dummies (Prior field of study) [a]				
Biomedical sciences				1.21 [0.76, 1.95]
Chemistry				0.83 [0.39, 1.80]
Liberal arts and sciences				1.13 [0.14, 9.39]
Medicine				0.54 [0.11, 2.77]
Pharmaceutics				2.57 [0.86, 7.72]
Psychology				1.04 [0.41, 2.64]
Other				0.72 [0.31, 1.66]
		Random effects		
Variance components				
Level 1	3.29	3.29	3.29	3.29
Level 2	0.24	0.25	0.24	0.21
Level 3	< .001	< .001	< .01	< .01
		Goodness of fit		
Deviance	3153.91	3148.05	3144.29	3138.16
Number of estimated parameters	3	4	5	12
Model comparison test		$\chi^2(1) = 5.87$*	$\chi^2(1) = 3.76$*	$\chi^2(7) = 6.17$
AIC	3159.91	3156.05	3154.29	3162.16

Note. The reported estimates of predictors are odds ratios. Confidence intervals are in square brackets.

[a] The reference category: "Biology".

*p < .05. **p < .01. ***p < .001.

4.3.2 Graduate Grade Point Average as a Dependent Variable

Table 4.1 depicts the results for the incremental validity of Type of prior HEI and field of study beyond percentile rank of UGPA. Model 2 shows that students from Dutch universities of applied sciences attainted significantly lower GGPAs compared to students from Dutch research universities, Dutch university colleges, and foreign HEIs. The addition of type of HEI increased the explained variance in GGPA by a small amount (2%). Adding field of study in Model 3 does not improve the model significantly. The

model with all study variables explained almost one-third of the total variance in GGPA.

Table 4.2 shows the incremental validity of prior thesis grade beyond percentile rank of UGPA (Model 2). Adding dummies on field of study did not improve the explanatory power of the model (Model 3). The model with significant predictors (Model 2) explained substantial amount of variance in GGPA (40%).

Table 4.1 | *Hierarchical Regression Results for Graduate Grade Point Average. Sample 3.*

Variable	Model 0	Model 1	Model 2	Model 3
		Fixed effects		
Intercept	7.81*** [7.69, 7.93]	7.82*** [7.74, 7.90]	7.61*** [7.49, 7.73]	7.56*** [7.36, 7.76]
Percentile rank of UGPA		0.01*** [0.01, 0.01]	0.01*** [0.01, 0.01]	0.01*** [0.01, 0.01]
Dummies (Type of prior HEI)[a]				
Dutch research university			0.23*** [0.13, 0.33]	0.25** [0.07, 0.43]
Dutch university college			0.29*** [0.15, 0.43]	0.39** [0.10, 0.68]
Foreign HEI			0.25*** [0.11, 0.39]	0.25* [0.05, 0.44]
Dummies (Prior field of study)[b]				
Biotechnology				0.15 [-0.05, 0.34]
Biology and medical laboratory				0.05 [-0.15, 0.25]
Biomedical sciences				0.02 [-0.06, 0.10]
Chemistry				0.01 [-0.11, 0.13]
Liberal arts and sciences				-0.06 [-0.28, 0.16]
Medicine				0.22 [-0.02, 0.46]
Pharmaceutics				0.15* [0.03, 0.27]
Psychology				-0.08 [-0.21, 0.05]
Other				0.03 [-0.08, 0.14]
		Random effects		
Variance components				
Level 1	0.27	0.20	0.20	0.20
Level 2	0.01	< 0.01	< 0.01	< 0.01
Level 3	0.04	0.02	0.01	0.02
Total explained variance (%)		30	32	32
		Goodness of fit		
Deviance	2463.46	1985.87	1960.12	1946.51
Number of estimated parameters	4	5	8	17
Model comparison test		$\chi^2(1) = 477.59$***	$\chi^2(3) = 25.75$***	$\chi^2(9) = 13.61$

Note. Confidence intervals are in square brackets. [a] The reference category: "Dutch university of applied sciences". [b] the reference category: "Biology".
*$p < .05$. **$p < .01$. ***$p < .001$.

4.3.3 Graduate Time to Degree as a Dependent Variable

Table 5.1 shows that the predictive validity of UGPA toward graduate time to degree holds even after accounting for the dependency of students within hierarchical structure: the higher the percentile rank of UGPA, the shorter graduate time to degree (Model 1). Adding Type of prior HEI significantly improves the model fit (Model 2).

Students, who completed their undergraduate degree outside of the Netherlands, have significantly shorter time to graduate degree than students from Dutch research universities and students from Dutch university colleges. Further, adding field of study above and beyond UGPA significantly improves the model fit again (Model 3). A dummy "psychology versus biology" is a significant predictor: Students who studied psychology during their bachelor's program take significantly more time to complete their research master's program than students who studied biology. The total explained variance of the model with all study variables included is very small.

Table 4.2 | *Hierarchical Regression Results for Graduate Grade Point Average. Sample 4.*

Variable	Model 0	Model 1	Model 2	Model 3
		Fixed effects		
Intercept	7.82*** [7.70, 7.94]	7.84*** [7.77, 7.93]	7.86*** [7.80, 7.91]	7.84*** [7.78, 7.80]
Percentile rank of prior average grade		0.01*** [0.01, 0.01]	0.01*** [0.01, 0.01]	0.01*** [0.01, 0.01]
Prior thesis grade			0.21*** [0.16, 0.25]	0.21*** [0.16, 0.25]
Dummies (Prior field of study)[a]				
BMS				0.04 [-0.04, 0.12]
Chemistry				-0.02 [-0.14, 0.10]
Liberal arts and sciences				-0.02 [-0.29, 0.25]
Medicine				0.10 [-0.19, 0.38]
Pharmaceutics				0.12 [-.01, 0.25]
Psychology				-0.02 [-0.16, 0.12]
Other				0.01 [-0.14, 0.15]
		Random effects		
Variance components				
Level 1	0.28	0.19	0.18	0.18
Level 2	0.01	< 0.01	< 0.01	< 0.01
Level 3	0.03	0.01	0.01	< 0.01
Total explained variance (%)		35	40	41
		Goodness of fit		
Deviance	1765.58	1348.2282.16	1272.64	1267.76
Number of estimated parameters	4	5	6	13
Model comparison test		$\chi^2(1) = 417.35$***	$\chi^2(1) = 75.58$***	$\chi^2(7) = 4.88$

Note. Confidence intervals are in square brackets. [a] The reference category: "Biology".
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 5.2 shows that undergraduate thesis grade has incremental predictive validity above and beyond UGPA: Students with higher undergraduate thesis grade take less time to complete a research master's program (Model 2). The addition of field of study to this model improves the model fit (Model 3). Students with undergraduate degrees, referred to as the "other backgrounds" category, take significantly longer to complete their degree than students with undergraduate degrees in biology. The amount of explained variance in graduate time to degree is small.

Table 5.1 | *Hierarchical Regression Results for Graduate Time to Degree. Sample 3.*

Variable	Model 0	Model 1	Model 2	Model 3
		Fixed effects		
Intercept	29.64*** [28.78, 30.50]	29.57*** [28.63, 30.51]	28.56*** [26.91, 30.21]	28.21*** [25.15, 31.27]
Percentile rank of UGPA		-0.04*** [-0.06, -0.02]	-0.04*** [-0.06, -0.02]	-0.04*** [-0.06, -0.02]
Dummies (Type of prior HEI)[a]				
Dutch research university			1.37 [-0.12, 2.86]	1.96 [-0.96, 4.88]
Dutch university college			0.62 [-1.73, 2.97]	0.36 [-4.40, 5.12]
Foreign HEI			-2.14 [-4.28, -0.00]	-1.94 [-5.15, 1.29]
Dummies (Prior field of study)[b]				
Biotechnology				-0.04 [-3.16, 3.08]
Biology and medical laboratory				0.61 [-2.82, 4.04]
BMS				-0.77 [-1.86, 0.33]
Chemistry				-1.20 [-3.06, 0.66]
Liberal arts and sciences				0.83 [-2.54, 4.20]
Medicine				2.32 [-1.48, 6.12]
Pharmaceutics				-0.83 [-2.81, 1.15]
Psychology				2.99** [0.89, 5.09]
Other				0.87 [-0.85, 2.59]
		Random effects		
Variance components				
Level 1	54.51	53.58	52.88	52.34
Level 2	1.06	0.80	0.85	0.83
Level 3	1.60	2.05	2.00	1.43
Total explained variance (%)		1	3	4
		Goodness of fit		
Deviance	10771.21	10742.23	10722.34	10703.77
Number of estimated parameters	4	5	8	17
Model comparison test		χ^2 (1) = 28.98***	χ^2 (3) = 19.89***	χ^2 (9) = 18.57*

Note. Confidence intervals are in square brackets.

[a] The reference category: "Dutch university of applied sciences". [b] the reference category: "Biology"

The analysis of other dummies in Model 2, also showed significant effects of dummy variables "Foreign HEI versus Dutch research university" (b = -3.51***, 95%CI [-5.16, -1.86]) and "Foreign HEI versus Dutch university college" (b = -2.76*, 95%CI [-5.19, -0.33]). For the dummy, "Foreign HEI versus Dutch research university", the direction of the effects and the significance levels of the same dummies hold in Model 3 (b = -3.90***, SE = 0.94, 95%CI [-5.74, -2.06]). For the dummy "Foreign HEI versus Dutch university college", there is no significant effect in Model 3 after adding field of study dummies in Model 3 (b = -2.3, 95%CI [-6.49, 1.89]).

*p < .05. **p < .01. ***p < .001.

Table 5.2 | *Hierarchical Regression Results for Graduate Time to Degree. Sample 4.*

Variable	Model 0	Model 1	Model 2	Model 3
		Fixed effects		
Intercept	30.41*** [29.53, 31.29]	30.35*** [29.39, 31.31]	30.29*** [29.33, 31.25]	30.50* [29.34, 31.66]
Percentile rank of prior average grade		-0.04*** [-0.06, -0.02]	-0.01[-0.03, 0.01]	-0.02 [-0.04, <.001]
Prior thesis grade			-1.98*** [1.16, 2.80]	-1.92*** [-2.72, 1.12]
Dummies (Prior field of study) [a]				
Biomedical sciences				-0.71 [-1.98, 0.56]
Chemistry				-2.10 [-4.24, 0.04]
Liberal arts and sciences				2.34 [-2.66, 7.34]
Medicine				2.74 [-2.32, 7.80]
Pharmaceutics				-1.11 [-3.34, 1.12]
Psychology				2.24 [-0.19, 4.67]
Other				2.74* [0.17, 5.31]
		Random effects		
Variance components				
Level 1	59.33	58.23	56.90	56.23
Level 2	1.06	0.88	1.13	1.02
Level 3	1.32	1.75	1.75	1.18
Total explained variance (%)		1	3	5
		Goodness of fit		
Deviance	7722.91	7702.21	7679.92	7662.95
Number of estimated parameters	4	5	6	13
Model comparison test		$\chi^2(1) = 20.70$***	$\chi^2(1) = 22.28$***	$\chi^2(7) = 16.97$*

Note. Confidence intervals are in square brackets. [a] The reference category: "Biology".
*$p < .05$. **$p < .01$. ***$p < .001$.

4.4 Discussion

We tested whether (and to what extent) we can predict graduate study success using student undergraduate academic indicators in a sample of students across several masters' programs in the life sciences. Our study found that the strongest predictor was percentile rank of UGPA which showed predictive validity toward all three outcomes: The higher percentile rank of UGPA was related to higher odds of completing a graduate program, higher GGPA, and shorter time to degree. Undergraduate thesis grade had incremental validity beyond UGPA in predicting GGPA and graduate time to degree: the higher undergraduate thesis grade, the higher GGPA and shorter graduate time to degree. Type of prior HEI was found to be predictive of degree completion and GGPA: Students from Dutch research universities and from foreign HEIs have higher odds of completing a graduate program compared to students from Dutch universities of applied sciences and students from Dutch university colleges. Finally, only a few dummies of field of study were found as significant predictors of graduate study success dimensions. We found that our models explain substantial amounts of variance in GGPA

but not in graduate time to degree. We also found that our models predicted odds of graduate degree completion to a small extent.

4.4.1 Predictive Value of Undergraduate Academic Indicators for Graduate Degree Completion

Our findings indicate that the examined undergraduate academic indicators (namely, UGPA and type of prior HEI) to a small extent predict graduate degree completion, as we hypothesized in Hypothesis 1. We see two possible complementary reasons for this finding. The first reason might be that the dropping out of students in this sample was not related to their academic ability but to other factors during their masters' programs. As the empirical research shows, these could be reasons related to psychological resources, personality, study motivation, study conditions, study decisions, institutional guidance, and study performance during a graduate program (Cox et al., 2009; Heublein, 2014). The second plausible reason might be that degree completion is determined by conscientiousness, motivation, drive, interest, or adaptability (Kuncel et al., 2014; Schwager et al., 2015) and, therefore, it is a hard-to-predict outcome, especially using prior academic indicators which do not directly assess these qualities. It might be that methods that evaluate noncognitive constructs (e.g., conscientiousness or time management; Butter & Born, 2012) or advanced assessment of academic work (presentations, various operationalizations of research experience; Pacheco et al., 2015) are better suited for prediction of degree completion.

4.4.2 Predictive Value of Undergraduate Academic Indicators for GGPA

When predicting GGPA, the strongest predictor in our analysis was UGPA. The predictive validity of UGPA corroborates our Hypothesis 1 which was based on previous studies showing a stable relation between UGPA and GGPA (Chadi & de Pinto, 2018; Howell et al., 2014; Park et al., 2018; Zimmermann et al., 2018). Undergraduate thesis grade showed incremental validity above UGPA and slightly improved the predictive power of our model. Considering that prior studies on undergraduate research experience as a predictor of GGPA have never operationalized it through undergraduate thesis grade (see Miller et al., 2021, for the overview of prior operationalizations), we cannot place our finding in the context of literature. However, it does align with our Hypothesis 2: We expected the undergraduate thesis grade to be positively related to graduate success because the undergraduate thesis is the most relevant undergraduate activity for research-oriented graduate studies. In general terms, it corroborates the meta-analytical findings which show that prior achievement (in this case, performance on a research-related task such as undergraduate thesis) is one of the best predictors of future achievement (Richardson et al., 2012; Schneider & Preckel, 2017). We consider that it could be beneficial to explore this operationalization further, especially as it allows us to place students on one metric, at least those who come from the same prior

HEI. In doing so, it is important to keep the possible effects of unintentional internal grading culture in mind.

Our next finding regarding prediction of GGPA is that students from universities of applied sciences obtain significantly lower GGPAs than students from other types of HEIs. This can be explained by a more practice-oriented curriculum of universities of applied sciences versus a research-oriented curriculum of research universities. It makes sense that the lack of preparation for the theoretical aspects of research places these students at a disadvantage compared with students from research oriented HEIs and leads at the end of a research master's program to lower GGPA. It is important to note, however, that despite incremental validity beyond and above UGPA, the gain in explained variance from type of HEI is small, in line with Hypothesis 3. This means that GGPA is not heavily determined by type of prior HEI in presence of UGPA, as we expected. As for field of study, the finding of the only significant dummy of field of study out of 10 and the negligible gain in amount of explained variance was not surprising and was expected in Hypothesis 4. Considering the examined masters' programs are interdisciplinary, their curriculum components are designed for students from different fields of studies.

4.4.3 Predictive Value of Undergraduate Academic Indicators for Graduate Time to Degree

We found that all four examined undergraduate academic indicators predict graduate time to degree only to a small extent. This result is in line with findings on undergraduate level where it was shown that precollege characteristics account for a small amount of variance in time to degree (Yue & Fu, 2017). We have two possible explanations for our finding. The first explanation is that among the undergraduate academic indicators we examined, there were none that measured motivation of students. There are, however, some indications that intrinsic motivation exerts positive influence on study progress (Slijper et al. 2016). We could not use assessments of motivation due to a practical reason (they were not available in our institutional data). We would also like to note that the existing selection methods based on motivation such as personal statements have not been shown as valid instruments (Murphy et al., 2009). Thus, we do not expect that having assessments of motivation available would deliver a substantial gain in explained variance in graduate time to degree.

Our second explanation is that what occurs during a graduate program plays a more important role in graduate study delays than undergraduate academic indicators. The factors during a graduate program that are influential for study delays are individual (e.g., student sense of belonging), supervisory (e.g., clarity of supervisor's communication toward their student), and departmental/institutional (e.g., graduate policies and practices, workload during a program; de Valero, 2001; Ruete et al., 2021; van de Schoot et al., 2013; van Rooij et al., 2021). In additon to these three factors, we think

that research masters' students might feel pressure to produce early-career publications (Crane & Pearson, 2011) because publishing academic work makes a difference when applying to a research-oriented position in the future (Stoilescu & McDougall, 2010). This pressure impacts students' decisions to produce a publication at the cost of longer time to degree. Overall, it appears that time to degree represents a variable that is hard to predict using information available upon admissions to a graduate program.

Our results regarding international students and time to degree are striking. Although the international students may experience a cultural shock (Zhou et al., 2008), both within and outside of studies (housing, teaching methods, and adjustment to new culture), they still take less time to complete a master's degree than local students who had research-intensive undergraduate education in their own country. Some of this can be explained by the fact that all international students in our analyses were from the EU, therefore, their academic and social integration scores are comparable to domestic students (Rienties et al., 2014). This finding might also be explained by the fact that many international students receive grants or loans as a part of an international exchange program, and this funding is usually provided for the official duration of their master's program (e.g., for two years in the case of masters' programs addressed in this study). Therefore, finishing on time could be a strong motivator for international students because it prevents them from taking out further loans or having to apply for additional grant money.

4.4.4 Theoretical Contributions

The aspects of undergraduate research work previously examined such as "undergraduate research experience present versus undergraduate research experience absent" (Cox et al., 2009; Miller et al., 2021) or "duration of research experience" (Gilmore et al., 2015; Hall et al., 2017; Weiner, 2014) do not reflect the *quality* of students' undergraduate research work. Our study, to the best of our knowledge, was the first to examine whether assessments regarding quality of undergraduate students' independent research work (i.e., a grade for undergraduate thesis) predict graduate study success. Thus, the promising results of our study provide another potentially effective direction in which researchers could operationalize undergraduate research experience in the future.

This study also added to our understanding whether type of prior HEI and prior field of study are reasonable predictors of graduate study success, as only a handful of studies exist on this topic (Chadi & de Pinto, 2017; Craig et al., 2004). By including students from Dutch universities of applied sciences, we were able to show (small) differences in their performance on two graduate study success dimensions (degree completion and GGPA) comparing to all other groups of students. Examining differential graduate study success of students from different types of HEIs and field of studies is important considering the problem of equal access to graduate education. It is, therefore, recommended that HEIs with research-intensive training, who wish to prioritize equal access to graduate

education, consider providing additional support for students with undergraduate degrees from non-research-intensive HEIs. It is also recommended to carefully review the extent of decisive power that is assigned to field of study as a selection method and possibly lessen it.

4.4.5 Practical Contributions

The application of undergraduate thesis grade as a selection method could be considered in practice, especially in programs with a similar research-oriented focus and where admission committees regard GGPA as an important dimension of graduate study success of their students. However, we call for a conscious choice in doing so. If we select students, who are already good in what they are supposed to do during their graduate program, what is the added value of the program in the learning process? Do we not exclude students who come from nonresearch undergraduate schools? Or should programs select and teach those who will gain the most (e.g., students who were less successful in research-related tasks such as an undergraduate thesis or simply never had a chance to work on a thesis during their undergraduate studies). We suggest that universities and graduate programs make this decision of using undergraduate thesis grade for selection purposes, accounting for their mission statements and vision of their student body.

Another practical consideration regarding the implementation of grades is that although this study showed their predictive validity (i.e., of undergraduate thesis grade and UGPA), it is important to account for the context in which these grades were obtained. While the traditional meritocratic equality of opportunity model of fair access implies that study places go to the most highly capable, irrespective of their social-economic background, an alternative model is gaining recognition and states that indicators of merit, including grades, need to be assessed contextually in light of an applicant's socioeconomic circumstances (Boliver & Powell, 2021). To make this possible, the individual HEIs should be allowed to gather and use data on socioeconomic status for conducting research on this topic which is almost never the case in certain European countries (partly a consequence of the recently adopted European Data Protection Regulation).

As for the other examined undergraduate academic indicators, the addition of type of prior HEI and prior field of study does not seem to deliver sufficient incremental validity once grades are in the model. Moreover, when applying prior HEI as a selection criterion, admissions committees should consider the indications that certain types of prior HEI are associated with lower socioeconomic status (e.g., on average, students from Dutch universities of applied sciences tend to have lower socioeconomic status than their peers in research universities; The Netherlands Association of Universities of Applied Sciences, 2012). Therefore, the application of type of prior HEI as a selection criterion could mask student selection based on socioeconomic characteristics which

would be morally and legally inappropriate. Instead, it might be practical to provide these students with additional guidance during their graduate studies to ensure graduate study success.

4.4.6 Limitations

This study does not come without limitations. First, we used data of already selected students and did not have information on how students, who were not selected, would have performed. However, since we were interested in detecting relationships between undergraduate academic indicators and graduate study success and not establishing the means or cut-off scores, there is no reason to assume that these relationships would be fundamentally different in a wider sample of all applicants.

The next limitation is that student admissions data registered in the administrative system at this graduate school are limited to variables from official transcripts. The scores on other documents that require additional assessment of admissions committees (recommendation letters, interviews, personal statements, etc.) were not standardized across programs at this graduate school, therefore, could not be included into the statistical analysis. However, the fact that our data came from official transcripts basically excluded the possibility of unreliable data. Moreover, the undergraduate academic indicators, which were the focus of this paper, are usually present in most similar graduate schools' data sets which allows considerable generalizability of our findings.

Another limitation is that this research is conducted within one graduate school of one university. However, students from thirteen different research masters' programs were included from relatively diverse field of studies which provides an opportunity for a certain generalizability of our findings for other research-oriented graduate programs.

Finally, it is important to note that our finding on limited predictive value of prior field of study should be regarded with a certain degree of caution because of the limited range of this variable in our sample. Although there were students with prior field of studies distinct from the life sciences, such as psychology, most came from disciplines closely related to the life sciences. Therefore, we leave a possibility open that our findings regarding this predictor could be different if we had data with a wider range of prior field of studies.

4.5 Conclusions

In this study, we aimed to validate certain widely used undergraduate indicators to help create a more objective, efficient, and inclusive master's admissions process. What we found is that undergraduate thesis grade is a valid predictor of GGPA in addition to UGPA. Therefore, these indicators should be considered for selection purposes for research oriented graduate programs in the life sciences and possibly for programs with

a similar focus. We also showed that type of prior HEI and prior field of study do not add much to the prediction of graduate study success after the prior grades have been taken into consideration. All examined undergraduate academic indicators did not contribute much to prediction of graduate degree completion and time to degree. While this study took place in a Dutch HEI, our findings, especially those on UGPA, undergraduate thesis grade, and field of study, are generalizable to research-intensive programs across EHEA. The graduate programs outside EHEA can consider them as well, accounting for the differences in structure of graduate programs. Likewise, our models, which combined different international student groups by using percentile ranks, can be applied across different HEIs.

4.6 References

AHZ Associates. (2021). *The difference between taught vs. research masters*. https://ahzassociates.co.uk/taught-masters-vs-research-masters/

Association of Universities in the Netherlands. (2021). *Instroom nieuwe masterstudenten [Influx of new masters' students].* https://vsnu.nl/instroom-master.html

Boliver, V., & Powell, M. (2021). *Fair admission to universities in England: Improving policy and practice*. Nuffield Foundation. https://www.nuffieldfoundation.org/wp-content/uploads/2021/01/Fair-admission-to-universities-in-England.pdf

Bridgeman, B., Burton, N., & Cline, F. (2009). A note on presenting what predictive validity numbers mean. *Applied Measurement in Education*, *22*(2), 109–119. https://doi.org/10.1080/08957340902754577

Burmeister, J., McSpadden, E., Rakowski, J., Nalichowski, A., Yudelev, M., & Snyder, M. (2014). Correlation of admissions statistics to graduate student success in medical physics. *Journal of Applied Clinical Medical Physics*, *15*(1), 375–385. https://doi.org/10.1120/jacmp.v15i1.4451

Burton, N. W., & Wang, M. (2005). *Predicting long-term success in graduate school: A collaborative validity study.* (Report No. 99-14R. ETS RR-05-03). Educational Testing Service. http://grad.uga.edu/wp-content/uploads/2017/09/GRE_Research_Report.pdf

Butter, R., & Born, M. P. (2012). Enhancing criterion-related validity through bottom-up contextualization of personality inventories: The construction of an ecological conscientiousness scale for Ph.D. candidates. *Human Performance*, *25*(4), 303–317. https://doi.org/10.1080/08959285.2012.703730

Chadi, A., & de Pinto, M. (2018). Selecting successful students? Undergraduate grades as an admission criterion. *Applied Economics*, *50*(28), 3089–3105. https://doi.org/10.1080/00036846.2017.1418072

Cox, G. W., Hughes, W. E., Jr, Etzkorn, L. H., & Weisskopf, M. E. (2009). Predicting computer science Ph.D. completion: A case study. *IEEE Transactions on Education*, *52*(1), 137–143. https://doi.org/10.1109/TE.2008.921458

Craig, P. L., Gordon, J. J., Clark, R. M., & Langendyk, V. (2004). Prior academic background and student performance in assessment in a graduate entry programme. *Medical Education*, *38*, 1164–1168. https://doi.org/10.1111/j.1365-2929.2004.02043.x

Crane, N. J., & Pearson, Z. (2011). Can we get a pub from this? Reflections on competition and the pressure to publish while in graduate school. *The Geographical Bulletin*, *52*, 77–80. https://gammathetaupsilon.org/the-geographical-bulletin/2010s/volume52-2/article4.pdf

Dabney, K. P. T. (2012). *Differences within: A comparative analysis of women in the physical sciences-motivation and background factors* (Publication No: AAT 3507074) [Doctoral dissertation, University of Virginia]. ProQuest Dissertations and Theses. https://doi.org/10.18130/V3XZ9R

Dore, M. L. (2017). *Factors in the admissions process influencing persistence in a Master's of Science program in marine science* [Doctoral dissertation, Nova Southeastern University]. https://nsuworks.nova.edu/fse_etd/110/

European Higher Education Area. (n.d.). *Three-cycle system*. https://www.ehea.info/page-three-cycle-system

Eurydice Network. (2020). *Key features of the Dutch education system*. https://eacea.ec.europa.eu/national-policies/eurydice/content/netherlands_en

Fu, Y. (2012). *The effectiveness of traditional admissions criteria in predicting college and graduate success for American and international students*. [Doctoral dissertation, University of Arizona]. ProQuest LLC. https://eric.ed.gov/?id=ED536665

Gilmore, J., Vieyra, M., Timmerman, B., Feldon, D., & Maher, M. (2015). The relationship between undergraduate research participation and subsequent research performance of early career STEM graduate students. *The Journal of Higher Education*, *86*(6), 834–863. https://doi.org/10.1353/jhe.2015.0031

Hall, J. D., O'Connell, A. B., & Cook, J. G. (2017). Predictors of student productivity in biomedical graduate school applications. *PLOS ONE*, *12*(1), Article e0169121. https://doi.org/10.1371/journal.pone.0169121

Heublein, U. (2014). Student drop-out from German higher education institutions. *European Journal of Education*, *49*(4), 497–513. https://doi.org/10.1111/ejed.12097

Howell, L. L., Sorenson, C. D., & Jones, M. R. (2014). Are undergraduate GPA and General GRE percentiles valid predictors of student performance in an engineering graduate program? *International Journal of Engineering Education*, *30*(5), 1145–1165. https://scholarsarchive.byu.edu/facpub/1343/

Kogan, L. R., Stewart, S. M., Schoenfeld-Tacher, R., & Hellyer, P. W. (2015). Perceptions of veterinary admissions committee members of undergraduate credits earned from community colleges or online compared to traditional 4-year institutions. *Open Veterinary Journal*, *5*(1), 71–84. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4629565/

Kuncel, N. R., Credé, M., & Thomas, L. L. (2007). A meta-analysis of the predictive validity of the Graduate Management Admission Test (GMAT) and undergraduate grade point average (UGPA) for graduate student academic performance. *Academy of Management Learning & Education*, *6*(1), 51–68. https://doi.org/10.5465/amle.2007.24401702

Kuncel, N. R., Kochevar, R. J., & Ones, D. S. (2014). A meta-analysis of letters of recommendation in college and graduate admissions: Reasons for hope. *International Journal of Selection and Assessment*, *22*(1), 101–107. https://doi.org/10.1111/ijsa.12060

Kurysheva, A., van Rijen, H. V. M., & Dilaver, G. (2019). How do admissions committees select? Do applicants know how they select? Selection criteria and transparency at a Dutch University. *Tertiary Education and Management*, *25*(4), 367–388. https://doi.org/10.1007/s11233-019-09050-z

Lub, A., van der Wende, M., & Witte, J. (2003). Bachelor-master programmes in the Netherlands and Germany. *Tertiary Education and Management*, *9*(4), 249–266. https://doi.org/10.1080/13583883.2003.9967108

MasterMind Europe. (2017). *Admissions to English-taught programs (ETPs) at master's level in Europe - procedures, regulations, success rates and challenges for diverse applicants*. https://mastermindeurope.eu/tools-and-reference-material/reports/

Miller, A., Crede, M., & Sotola, L. K. (2021). Should research experience be used for selection into graduate school: A discussion and meta-analytic synthesis of the available evidence. *International Journal of Selection and Assessment*, *29*(1), 19–28. https://doi.org/10.1111/ijsa.12312

Moneta-Koehler, L., Brown, A. M., Petrie, K. A., Evans, B. J., & Chalkley, R. (2017). The limitations of the GRE in predicting success in biomedical graduate school. *PLOS ONE*, *12*(1), Article e0166742. https://doi.org/10.1371/journal.pone.0166742

Murphy, S. C., Klieger, D. M., Borneman, M. J., & Kuncel, N. R. (2009). The predictive power of personal statements in admissions: A meta-analysis and cautionary tale. *College & University*, *84*(4), 83–88.

NVAO. (2016). *Specification of additional criteria for research master's programmes*. https://www.nvao.net/files/attachments/.1255/Nadere_uitwerking_aanvullende_criteria_onderzoeksmasters_30_mei_2016_EN_def_nieuwe_huisstijl.pdf

OECD. (2022). *International student mobility (indicator)*. https://doi.org/10.1787/4bcf6fc3-en

Pacheco, W. I., Noel, R. J., Porter, J. T., & Appleyard, C. B. (2015). Beyond the GRE: Using a composite score to predict the success of Puerto Rican students in a biomedical PhD program. *CBE Life Sciences Education*, *14*, 14:ar13, 1-14:ar13, 7. https://doi.org/10.1187/cbe.14-11-0216

Park, H.-Y., Berkowitz, O., Symes, K., & Dasgupta, S. (2018). The art and science of selecting graduate students in the biomedical sciences: Performance in doctoral study of the foundational sciences. *PLOS ONE*, *13*(4), Article e0193901. https://doi.org/10.1371/journal.pone.0193901

Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., & Cleland, J. (2016). How effective are selection methods in medical education? A systematic review. *Medical Education*, *50*(1), 36–60. https://doi.org/10.1111/medu.12817

Patterson, F., Roberts, C., Hanson, M. D., Hampe, W., Eva, K., Ponnamperuma, G., Magzoub, M., Tekian, A., & Cleland, J. (2018). 2018 Ottawa consensus statement: Selection and recruitment to the healthcare professions. *Medical Teacher*, *40*(11), 1–9. https://doi.org/10.1080/0142159X.2018.1498589

Posselt, J. R. (2016). *Inside graduate admissions: Merit, diversity, and faculty gatekeeping*. Harvard University Press. https://doi.org/10.1080/00220671.2016.1184524

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353–387. https://doi.org/10.1037/a0026838

Rienties, B., Luchoomun, D., & Tempelaar, D. (2014). Academic and social integration of master students: A cross-institutional comparison between Dutch and international students. *Innovations in Education and Teaching International*, *51*(2), 130–141. https://doi.org//10.1080/14703297.2013.771973

Ruete, D., Zavala, G., Leal, D., Lopez Lira, P., Medina, L. P. S. M., Aravena, M. E., & Costa, G. (2021, July 26-29). *Early detection of delayed graduation in master's students* [Paper presentation]. American Society for Engineering Education Annual Conference, Virtual Meeting. https://strategy.asee.org/36999

Salmi, J., & Bassett, R. M. (2014). The equity imperative in tertiary education: Promoting fairness and efficiency. *International Review of Education*, *60*(3), 361–377. https://doi.org/10.1007/s11159-013-9391-z

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin*, *143*(6), 565–600. https://doi.org/10.1037/bul0000098

Schwager, I. T. L., Hülsheger, U. R., Bridgeman, B., & Lang, J. W. B. (2015). Graduate student selection: Graduate record examination, socioeconomic status, and undergraduate grade point average as predictors of study success in a western European university. *International Journal of Selection and Assessment, 23*(1), 71–79. https://doi.org/10.1111/ijsa.12096

Snijder, J. K. (2016). *Mastering research: A study about the effectiveness of research master programmes in the Netherlands*. [Doctoral dissertation, University of Twente]. https://doi.org/10.3990/1.9789036541466

Statista. (2020). *Number of master's degrees earned in the United States from 1949/50 to 2029/30, by gender*. https://www.statista.com/statistics/185160/number-of-masters-degrees-by-gender-since-1950/

Stoilescu, D., & McDougall, D. (2010). Starting to publish academic research as a doctoral student. *International Journal of Doctoral Studies*, *5*, 79–92. http://ijds.org/Volume5/IJDSv5p079-092Stoilescu299.pdf

the Higher Education Statistics Agency. (2020). *Higher education student statistics: UK, 2018/19 - qualifications achieved*. https://www.hesa.ac.uk/news/16-01-2020/sb255-higher-education-student-statistics/qualifications

The Netherlands Association of Universities of Applied Sciences. (2012). *Achtergrondnotitie. Toegankelijkheid Hoger Onderwijs en de Rol Studiefinanciering [Background note. Access to higher education and the role of Student Finance].* https://www.vereniginghogescholen.nl/system/knowledge_base/attachments/files/000/000/390/original/Toegankelijkheid_hoger_onderwijs_rol_studiefinanciering_-_achtergrondnotitie_HBO-raad_over_leenstelsel.pdf?1443427985

The Quality Assurance Agency for Higher Education. (2020). *Characteristics statement: master's degree.* https://www.qaa.ac.uk/docs/qaa/quality-code/master%27s-degree-characteristics-statement.pdf?sfvrsn=86c5ca81_18

Torgerson, P. C., Gascoine, L., Heaps, C., Menzies, V., & Younger, K. (2014). *Higher education access: Evidence of effectiveness of university access strategies and approaches. A report to The Sutton Trust*. Durham University. https://www.suttontrust.com/wp-content/uploads/2019/12/Higher-Education-Access-Report-1.pdf

Torotcoi, S., Gologan, D., & Kurysheva, A. (2020). What works for underrepresented groups? Identifying effective measures for enhancing access to higher education in Europe. In A. Curaj, L. Deca, & R. Pricopie (Eds.), *European Higher Education Area: Challenges for a new decade* (pp. 177–196). Springer International Publishing. https://doi.org/10.1007/978-3-030-56316-5_13

van de Schoot, R., Yerkes, M. A., Mouw, J. M., & Sonneveld, H. (2013). What took them so long? Explaining PhD delays among doctoral candidates. *PLOS ONE*, *8*(7), Article e68839. https://doi.org/10.1371/journal.pone.0068839

van der Wende, M. (2020). Netherlands, higher education in. In M. David, & M. Amey (Eds.), *The SAGE encyclopedia of higher education* (Vol. 1, pp. 1079-1080). SAGE Publications, Inc., https://dx.doi.org/10.4135/9781529714395.n403

van Rooij, E., Fokkens-Bruinsma, M., & Jansen, E. (2021). Factors that influence PhD candidates' success: The importance of PhD project characteristics. *Studies in Continuing Education*, *43*(1), 48–67. https://doi.org/10.1080/0158037X.2019.1652158

Walpole, M., Burton, N. W., Kanyi, K., & Jackenthal, A. (2002). *Selecting successful graduate students: In-depth interviews with GRE users* (Report No. GREB-99-11R, RR-02-08). Educational Testing Service. https://doi.org/10.1002/j.2333-8504.2002.tb01875.x

Weiner, O. D. (2014). How should we be selecting our graduate students? *Molecular Biology of the Cell*, *25*(4), 429–430. https://doi.org/10.1091/mbc.e13-11-0646

Witte, J., van der Wende, M., & Huisman, J. (2008). Blurring boundaries: How the Bologna process changes the relationship between university and non-university higher education in Germany, the Netherlands and France. *Studies in Higher Education*, *33*(3), 217–231. https://doi.org/10.1080/03075070802049129

Wollast, R., Boudrenghien, G., Van der Linden, N., Galand, B., Roland, N., Devos, C., De Clercq, M., Klein, O., Azzi, A., & Frenay, M. (2018). Who are the doctoral students who drop out? Factors associated with the rate of doctoral degree completion in universities. *International Journal of Higher Education*, *7*(4), 143-156. https://doi.org/10.5430/ijhe.v7n4p143

Younger, K., Gascoine, L., Menzies, V., & Torgerson, C. (2018). A systematic review of evidence on the effectiveness of interventions and strategies for widening participation in higher education. *Journal of Further and Higher Education*, *43*(6), 1-32. https://doi.org/10.1080/0309877X.2017.1404558

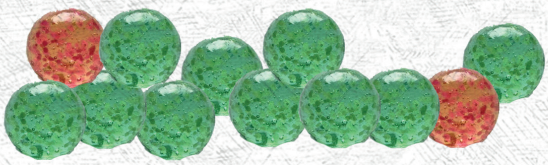
Yue, H., & Fu, X. (2017). Rethinking graduation and time to degree: A fresh perspective. *Research in Higher Education*, *58*(2), 184–213. https://doi.org/10.1007/s11162-016-9420-4

Zhou, Y., Jundal-Snape, D., Topping, K., & Todman, J. (2008). Theoretical models of culture shock and adaptation in international students in higher education. *Studies in Higher Education*, *33*(1), 63–75. https://doi.org/10.1080/03075070701794833

Zimmermann, J., Brodersen, K. H., Heinimann, H. R., & Buhmann, J. M. (2015). A model-based approach to predicting graduate-level performance using indicators of undergraduate-level performance. *Journal of Educational Data Mining*, *7*(3), 151–176. https://doi.org/10.5281/ZENODO.3554733

Zimmermann, J., von Davier, A. A., Buhmann, J. M., & Heinimann, H. R. (2017). Validity of GRE General test scores and TOEFL scores for graduate admission to a technical university in Western Europe. *European Journal of Engineering Education*, *43*(1), 144–165. https://doi.org/10.1080/03043797.2017.1343277

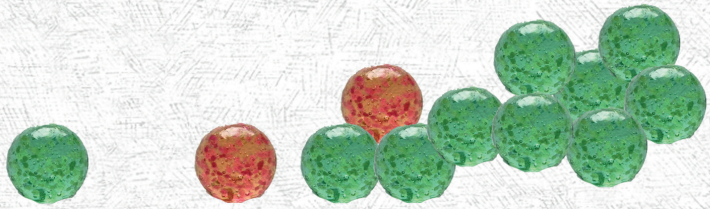
5



Acknowledgements

The authors are grateful to the Education Testing Service and its employees Jacqueline Briel, Brent Bridgeman, Harrison Kell, Fred Cline, Maria Smith, Maria-Victoria Calabrese, Stella Saliari, and Wouter Schut for their help in organizing this study. We would also like to thank Michele Gerbrands, Marjan Batist de Vos, Claudia Tielmans, Farah Kools, Ouafa Dahri, Pauline Krijgsheld, Rianne Bouwmeester, Dimitra Mousa, Renske de Kleijn, Nivard Koning, Anne-Petra Rozendal, Els van der Vlist, Krijn Vrijsen and Spiros Pachis for their help with managing the registration of students before the GRE testing event and proctoring during the event itself. We appreciate the assistance of Pauline Krijgsheld, Elena Donner, Yolande Lutteke, Timon de Boer, Shirrinka Goubitz, and Monica van der Garde in informing the students about this study. We are also grateful to Christine Fox for substantive editing of the text.

To Test or not to Test? The Graduate Record Examinations: Predictive Validity toward Graduate Study Success on Research Masters' Programs in a Large European University



Author note

An adapted version of this chapter is in preparation for submission to a peer-reviewed journal as "Kurysheva A., van Rijen H. V. M., & Dilaver G. (2022). *To Test or not to Test? The Graduate Record Examinations: Predictive Validity toward Graduate Study Success on Research Masters' Programs in a Large European University* [Manuscript in preparation]. BMS Department, Educational Center University Medical Center Utrecht".

All authors made substantial contributions to the study conception and design. A. Kurysheva overviewed the collection of the data, analyzed it, and prepared the draft of the manuscript. All authors provided critical feedback on the manuscript.

Abstract

Graduate admissions committees in Europe have a challenging task of selecting students from an increasingly large pool of candidates with diverse application files, as a result of growing cross-border mobility in the European Higher Education Area (EHEA). Graduate standardized testing is a tool that can ease the comparison of application files and potentially assists admissions committees in their decision-making. The purpose of this study was to examine whether the Graduate Record Examinations (GRE) is predictive of several dimensions of graduate success on English-taught research masters' programs in Science, Technology, Engineering, and Mathematics at a large European university. Hierarchical regression analyses were conducted. It was found that all GRE scales predicted Graduate Grade Point Average. Individual GRE scales predicted internship grade and supervisors' assessments of students' skills on performing research, their practical skills and structure and style of their research reports. None of the GRE scales predicted supervisors' assessments of content of students' research reports. These relationships held after accounting for socioeconomic status. Overall, the GRE appeared as a reasonable predictor of graduate study success. The pros and cons of the GRE implementation in admissions practices on European masters' programs are discussed. Moreover, benefits of using the GRE for certain groups of applicants as well as the legal limitations are outlined.

Keywords: Graduate Record Examinations, GRE, graduate study success, European university, research masters' programs, SES.

5.1 Introduction

In recent years, European higher education has seen a number of reforms. The most overarching, comprehensive, and long-lasting of these was the Bologna Process, which was introduced in 1999 to help strengthen the competitiveness and attractiveness of European higher education. The Bologna Process helped initiate the formation of the European Higher Education Area (EHEA) which aims to ensure comparability and compatibility of higher education degrees across 49 member countries (Bologna Working Group on Qualifications Frameworks, 2005). One of the main achievements of the EHEA has been the harmonization of frameworks for qualifications through the introduction of the three cycle (bachelor's/master's/PhD) degree structure. For higher education institutions (HEIs), this harmonization initiated three changes to the admissions process for graduate school applicants.

The first change was that students could transfer more easily between undergraduate (bachelor's) and graduate (master's and PhD) education not only within their own countries, but also between countries. While this change increased student mobility across EHEA member countries, this mobility is not being evenly distributed. The majority of EHEA countries have seen more outgoing than incoming students (Skinner, 2018). A smaller number of countries have experienced a steady increase of incoming international students. According to the most recent Bologna Process Implementation Report (European Commission et al., 2020), the Netherlands belongs to the top three EHEA countries, where incoming international students substantially prevail over outgoing students (with the UK and Denmark being the other two; see Figure 1).

The second admissions-related change was that it became easier for students to pursue graduate degrees in related fields, but not necessarily the same field as their undergraduate studies (Payne, 2015). For example, graduate programs in the life sciences started receiving applications from students with bachelors' degrees in areas such as psychology and mathematics. Finally, the third admissions-related change was that in countries with binary higher education systems, such as the Netherlands, students from universities of applied sciences[15] (which focus on higher professional education) were now allowed to directly apply to graduate programs at research universities (which focus on research-intense education).

While there are many benefits to these Bologna-related policy changes, there is also increased complexity due to diversity of admissions files and the absence of one common metric of comparison. The diversity of admissions files and the fact that in many HEIs the number of admissions exceeds the number of seats has resulted in a process of selection. This process of selection has two main purposes: (1) to determine which applicants are eligible in principle and (2) to rank applicants on their predicted success on a specific graduate program.

15 See Chapter 1 (General Introduction) for a detailed explanation of the pathways within the Dutch higher education system.

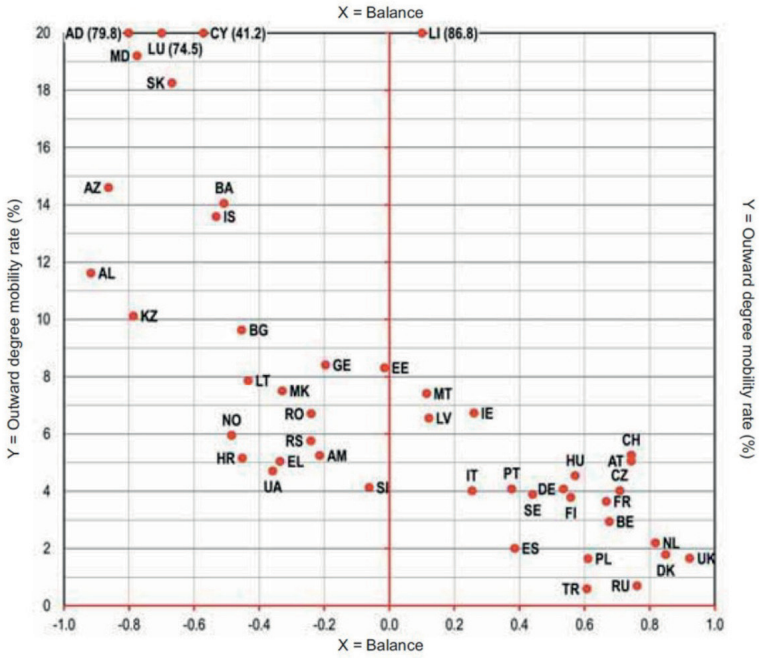


Figure 1 | *The Degree Mobility Balance in 2017 with Reference to the Outward Degree Mobility Rate of the Respective Country.*

© The European Higher Education Area in 2020 Bologna Process Implementation Report

Note. "Both axes include mobility flows within and outside the EHEA: The higher the importing balance, the lesser the outward mobility rate. For graphical readability purpose, balance is computed as the absolute difference (incoming – outgoing students) divided by the total number of incoming students (when the balance is positive) or by the total number of outgoing students (in case of negative balance). This results in a smoother continuum, more readable when plotted than taking the ratio (incoming/outgoing) which is below 1 for most countries." (Bologna Process Implementation Report, 2020, p.142). "The United Kingdom, Denmark, and the Netherlands are situated on the right side of the X-axis with the highest imbalance (above 82 % each) and very low shares of outgoing mobile students (below 2.5 %)" (Bologna Process Implementation Report, 2020, p. 143).

The complexity behind the student selection process, however, makes it challenging for admissions committees to understand and compare education systems and grading scales of other countries, especially of those outside of the EU. Within the EU, these comparisons are facilitated by the knowledge gained during the long-time investments in credit and grade transfer systems under the ERASMUS program—the EU's program to support, among others, mobility in higher education in Europe (European Commission, n.d.). However, even within the EU, the comparisons of education systems and grading scales are not always straightforward and are more complicated to conduct than within one country.

The admissions committees might also not understand the quality of education at an applicant's undergraduate HEI. For example, admissions committees, using international rankings to judge on quality of universities, can misinterpret those rankings because perceived quality factors do not usually align with methodologies

used by the ranking committees (Easley et al., 2021). Likewise, admissions committees might find it challenging to compare learning outcomes of applicants' programs. Attempts for such comparisons were shown to be susceptible to assessing only the differences in knowledge, while a student's learning outcomes are greatly conditioned by the profession and discipline (Caspersen et al., 2014).

One possible solution to ensure comparability of graduate school applications would be to introduce standardized admissions testing. Standardized admissions testing provides admissions committees with a metric for assessing students from diverse educational backgrounds. Therefore, they can be used to help create a fairer, more reliable, and transparent admissions process. Internationally, the most widely used standardized test is the Graduate Record Examination (GRE).

5.1.1 GRE

5.1.1.1 The GRE History

The GRE emerged from a collaborative project on studying outcomes of college education, which was run in the1930-s by the Carnegie Foundation for the Advancement of Teaching and four Ivy League universities (Columbia, Harvard, Princeton, and Yale; Altman et al., 1977). As its popularity grew, it was eventually taken over in 1948 by the recently founded Education Testing Service (ETS). Over the years, the GRE had differentiated into two types of tests: a test that measures general skills, which are not related to a specific field of study, such as critical thinking, analytical writing, quantitative and verbal reasoning (GRE General) and achievement tests that measure knowledge in a particular field of study (GRE Subject Tests in Psychology, Physics, Mathematics, Chemistry). This chapter is dedicated to the GRE General and does not address the GRE Subject Tests.

5.1.1.2 The GRE Structure

Since its inception, the GRE General has undergone several additions, including a significant revision in 2011 (Klieger et al., 2017). Presently, the GRE General measures Verbal Reasoning (section GRE-V), Quantitative Reasoning (section GRE-Q), and Analytical Writing (section GRE-A). Graduate school applicants must take all three sections of the GRE General (it is not possible to take one section of the test). For this study, we will use the terms GRE General and GRE interchangeably.

5.1.2 GRE Usage in the US and Europe

In the US, testing is firmly rooted in the education system, education regulations, and expectations of stakeholders (Payne, 2015). This has contributed to the widespread test taking of the GRE in the US and beyond. In 2015–2016, a milestone was reached worldwide when 584,677 GRE tests were taken (61.9% of those were in the US;

Educational Testing Service [ETS], 2019). These numbers have gradually decreased over the years: In 2018-2019, a total of 532, 826 tests were taken (55.5% of those in the US; ETS, 2021)[16]. This decrease reflects a recent turn away from standardized testing as an increasing number of PhD programs across Science, Technology, Engineering, and Mathematics (STEM) fields have dropped[17] their GRE requirement in admissions (King et al., 2020).

External admissions testing may not be a wide-spread tradition in Europe. The most European graduate schools base their admissions decisions primarily on prior grades and other admissions tools indicative of a student's merits. There are, however, several graduate schools that accept the GRE. Some European business schools have developed admissions procedures similar to the US model; see the list of European business programs that accept GRE scores: https://www.ets.org/gre/consider/business/programs). There are also a handful of masters' programs in other fields of studies that accept the GRE. For example, in the Netherlands, these are masters' programs in economics (e.g., from Utrecht University or Tilburg University), engineering (TU Delft), mathematics (University of Twente), and health sciences (Maastricht University) which either require or make the GRE General[18] optional for students with international bachelors' degrees. Despite these few exceptions, the GRE is not normally used for STEM programs in the Netherlands (and in EU countries in general). This is reflected in the open data of the ETS' where only 2.6% of the GRE test takers in 2015–2016 were from Europe and remained roughly the same for 2019–2020 (2.9%; ETS, 2019). While the GRE has been broadly accepted by graduate schools in the US (and, to a smaller extent, beyond the US) there are several questions regarding its usage. The most prominent questions are: (1) whether the GRE is a valid indicator of study success for STEM disciplines and (2) bias in standardized testing.

5.1.2.1 GRE as a Predictor of Graduate Study Success: Review of Literature

The empirical evidence shows that the GRE is a good predictor of first-year graduate grade point average (GGPA; Bridgeman et al., 2009; Burmeister et al., 2014; Fu, 2012; Moneta-Koehler et al., 2017), total GGPA (Burton & Wang, 2005; Howell et al., 2014; Klieger et al., 2014; Moneta-Koehler et al., 2017; Perez, 2011; Zimmermann et al., 2017), and faculty

16 During the covid-19 pandemic, the numbers of tests dropped even more: In 2019–2020, a total of 467,277 tests were taken (62.9% of those in the US; ETS, 2021); in 2020-2021, a total of 366,686 tests were taken (49,2% of those in the US; ETS, 2021).

17 Notably, this trend is not specific for admissions on a graduate level. Some leading universities such as University of California (UC) decided to suspend undergraduate standardized testing (American College Testing (ACT)/ Scholastic Aptitude Test (SAT)) until fall 2024 - the time needed to develop a new undergraduate test that better aligns with the UC's expectations for incoming students (UC Office of the President, 2020). Also, the Covid-19 pandemic has had a continuous impact on access to testing for applicants. Therefore, some elite higher education institutions have decided to allow applicants to submit their applications without requiring standardized test scores (e.g., Harvard College made such a decision for classes of '27, '28, '29, and '30).

18 As an alternative to the GRE General, these programs usually also accept score reports on the Graduate Management Admission Test (GMAT).

ratings of a student's academic and professional skills[19] (Burmeister et al., 2014; Burton & Wang, 2005; Howell et al., 2014; Moneta-Koehler et al., 2017). This is also confirmed in the meta-analytical studies (Kuncel et al., 2010; Kuncel & Hezlett, 2007, 2010).

As for other dimensions of graduate study success, scientific research has not provided a conclusive answer. For example, the primary empirical studies find no predictive validity of the GRE toward degree completion (Cox et al., 2009; Dore, 2017; Lorden et al., 2011; Lott et al., 2009; Miller et al., 2019; Moneta-Koehler et al., 2017; Petersen et al., 2018) and research productivity (Hall et al., 2017; Moneta-Koehler et al., 2017; Sanford, 2009; Sealy et al., 2019). However, the meta-analyses, which correct for statistical artifacts of primary studies (e.g., the restriction of range of a predictor), indicate a small in size, but positive significant relationship between the GRE and degree completion and research productivity (Kuncel & Hezlett, 2007a, 2007b, 2010). There are also multidirectional findings regarding the GRE as a predictor of graduate time to degree. While some authors have not found any predictive power (Dore, 2017; Hall et al., 2017; Moneta-Koehler et al., 2017; Petersen et al., 2018; Sealy et al., 2019), others have found a negative relationship (Howell et al., 2014; Lorden et al., 2011) and a positive relationship (Frasier, 2013).

Problem Statement Regarding Predictive Validity of GRE. The overwhelming majority of studies, mentioned above, were conducted within the US and their conclusions might not be directly transferable to the European context because of differences in structure of higher education systems. If we consider the scientific evidence available on the research-aspects of graduate study success within the US higher education context, we find that those studies examined only *productivity* aspects of graduate students' research work such as number of publications, conference presentations, awards, grants, and fellowships (Hall et al., 2017; Howell et al., 2014; Moneta-Koehler et al., 2017; Sanford, 2009; Sealy et al., 2019). It can be assumed that this focus on productivity relates to the fact that in the US, master's education is often integrated into overarching graduate programs that award students with a PhD degree at the end of their studies. PhD students obviously have higher chances to have research output published or presented. This is not necessarily the case in the European higher education context and its distinct master's cycle (i.e., master's as the final degree).

In European countries, to the best of our knowledge, only two studies on validity of the GRE exist (Schwager et al., 2015; Zimmermann et al., 2017). They both found that GRE is predictive for graduate GGPA. In addition to that, Schwager et al. (2015) found no relationship between GRE and degree completion, while Zimmermann et al. (2017) detected a weak positive relationship between GRE and study progress, but no relationship between GRE and thesis performance (defined as grade obtained for their master's thesis).

19 More specifically, faculty ratings mean that several faculty members (usually, two), who are familiar with a student, rate the student on "(a) professional knowledge, ability to apply that knowledge, and ability to learn independently (mastery of the discipline); (b) judgment in choosing professional issues and creativity and persistence in solving the issues (professional productivity); and (c) ability to communicate what was learned (communications skills)" (Burton & Wang, 2014)

The study of Zimmermann et al. (2017), which examined thesis performance, might be regarded as the first attempt to examine *quality* of graduate students' research work. The authors explain the finding on absence of relationship between the GRE and thesis performance by the differences in the grading schemes for theses, used by supervisors (Zimmermann et al., 2017). This leaves an open question regarding predictive validity of GRE toward students' research skills at a graduate level.

5.1.2.2 Bias in Standardized Testing. GRE and SES

Another consideration is whether standardized tests discriminate against students from certain groups, including racial/ethnic minorities, gender groups, and lower SES. Regarding racial inequality the US context, there is evidence that applicants from underrepresented groups (African American, Puerto-Rican, Native American, Mexican-American) obtain lower GRE scores than White and Asian American applicants as well as evidence that females obtain lower GRE scores than males (Miller & Stassun, 2014). These discrepancies are one of the reasons why the ETS recommends using multiple sources of information when making admissions decisions (ETS, n.d. -a)

Racial/ethnic minorities and females might also be susceptible to stereotype threat: "a psychological phenomenon in which a member of a negatively stereotyped group underperforms on an activity because of increased anxiety that they may confirm the negative stereotype" (Gordon, 2019, p. 387). However, the studies which examined this question show that negative effect of stereotype threat on performance of minorities occur when the instructions or subtle cues are embedded in the instructions before taking a test (Jamieson & Harkins, 2009; Schmader & Johns, 2003; Steele & Aronson, 1995). The cues which evoke stereotype threat might also arise from testing environment such as test-center size, their activity level, and formality level (warm/friendly as opposed to formal/professional) (Walters et al., 2004). In this regard, changing the practices that trigger stereotypes is seen as a primary solution to mitigate the effects of stereotype threat (Gordon, 2019).

Regarding SES, critics claim that students from families with higher SES achieve higher scores on standardized tests such as GRE on account of their access to professional test training and having more test experience (e.g., Clayton, 2016; Zwick, 2004). They also argue that standardized testing brings the financial burden on economically disadvantaged students (Ledford et al., 2020). An experimental study showed that low SES participants perform worse than high SES participants on GRE-like and IQ-like tests if a test is described as a measure of intellectual capability, but the performance of these two groups does not differ if they are presented with a test as nondiagnostic of intellectual capability (Croizet & Dutrevis, 2004).

Some researchers regard the weaker performance of low SES students on standardized tests not as a problem of standardized testing itself, but as a societal issue (the lack of financial and other resources puts them at a disadvantage). Therefore,

differential performance of socioeconomic groups on standardized tests, such as the GRE, simply reflects these existing societal inequalities (Kuncel & Hezlett, 2010). The meta-analytical research on standardized testing at the undergraduate level shows that even though SES is related to performance, the statistical control for SES reduces the estimated correlations between standardized tests and study success only to a small extent (Sackett et al., 2009). This means, according to Sackett et al. (2009), that the relationship between standardized admissions tests and study success is not an artifact of SES. There is no comparable meta-analysis on this topic on a graduate level. However, there is some evidence from primary studies, which points to the same conclusion (Schwager et al., 2015; Stricker & Rock, 1993).

Problem Statement Regarding GRE and SES. In order to sustain statistical power and the focus of our study, our research concentrates on whether SES affects the results on students' GRE scores (and not ethnicity or gender). Out of the two studies conducted on the GRE in the European context, only one included SES in the model. This study found that the detected significant relationship between GRE and GGPA is independent from SES (Schwager et al., 2015). Despite this result and the meta-analytical evidence on undergraduate level, it is important to account for SES while the debate surrounding the relationship between standardized test scores, SES, and study success is still ongoing (Rodríguez-Hernández et al., 2020).

5.1.2.3 Current Study

To enlarge the available scientific evidence on the GRE in the European context, specifically its value on English-taught STEM research masters' programs, we conducted a quantitative study, which used regression analysis. We also visualized the findings on predictive validity, showing the improvement in prediction (or its absence) once the GRE is added to the model.

We first examined whether the GRE is predictive above and beyond Undergraduate Grade Point Average **(**UGPA) toward several dimensions of graduate study success on research masters' programs. Next, we examined whether this relationship between the GRE and graduate study success still holds after SES is accounted for. Our focus on research masters' means that in addition to widely examined GGPA as one of the dimensions of graduate study success, this study also considers other, research-related, dimensions of graduate study success, such as assessments of the quality of independent research work of graduate students by their supervisors.

5.2 Method

5.2.1 Design and Procedure of the Study

The study was approved by the Netherlands Association for Medical Education Ethical Review Board (dossier number: 2018.5.9). One hundred and sixty-seven students from a large Dutch research university volunteered to participate in the study. These students started their studies in 2019 at one of three STEM-focused graduate schools (the Graduate School of Life Sciences (GSLS), the Graduate School of Natural Sciences (GSNS), and the Graduate School of Geosciences (GSGS). As a compensation for their time, they were offered a free lunch and a professional photograph for their LinkedIn profile.

The data were collected across two time points. Time point 1 was at the beginning of their two-year research master's program. The students, who had already been admitted and were thus enrolled in the programs, filled in a consent form, completed a paper-based GRE and a questionnaire on their background information during a 4-hour session. Due to the ongoing covid-19 pandemic, most of the students have not completed their research masters' programs within the two years they have been enrolled. To account for the delay, caused by the pandemic, we gathered the available data on graduate study success at 26 months after the start of their masters' programs and not after the traditional 24 months (Time point 2). Even after the two-month extension, 54% of the 167 students have still not completed their masters' degrees.

We did not collect data on "graduate time to degree", because the duration of studies of these students would not be comparable to a typical duration that one would have observed before the pandemic. Likewise, we did not collect data on "degree attainment", because many students experienced a significant delay in studies and perhaps still aim to finish their degrees. The remaining dimensions of graduate study success, which we considered, as well as predictors and control variables, are described in the "measures" section below.

5.2.2 The Institutional Context

The three graduate schools that participated in this study offer research-oriented English-taught education at a master's level (two-year programs of 120 ECTS credits in total) and a PhD level (four-year programs). For this study, we focus on the master's level programs. The demand for study placement at these graduate schools increases annually. For example, there were over 1,700 applicants with complete dossier in the 2020-2021 academic year at the GSLS, yet there were only 500–550 student places. The application files are diverse in these graduate schools. For example, almost half of the GSLS applicants (44%) were international students in the 2020–2021 academic year. The selection procedure is as follows. First, eligibility of applicants is established based on undergraduate degree. For international students, the university international

admissions office assesses the extent, to which their undergraduate degree corresponds to a Dutch undergraduate degree. Second, the applicants are ranked according to several aspects of their undergraduate degree such as quality of undergraduate HEI (based on curriculum and/or international university rankings) and study success during undergraduate degree, including UGPA, grades for undergraduate courses, relevant to the content of a master's program, as well as their motivation, extracurricular activities, and recommendations from referees. It is important to note that GRE scores are not considered during the selection process to these programs.

Once enrolled, the students receive intense research training during courses and internships. At the GSNS and GSGS, the majority of ECTS credits are earned through courses (compulsory and elective), while internships constitute a smaller part of the curriculum (typically, around 45 ECTS credits at the GSNS and 22.5–60 ECTS credits at the GSGS). At the GSLS, in contrast, the majority of ECTS credits are earned through internships (so-called major research project and minor research project or a research profile of 51 and 33 ECTS credits, respectively), while courses constitute a smaller part of the curriculum (usually, 27 ECTS credits). Nevertheless, students are recommended to take mandatory courses of at least 15 ECTS credits before taking an internship. The GSLS is also distinct from the other two graduate schools in terms of using specific rubrics for assessment of different aspects of students' performance during internships which will be explained in detail in the section below.

5.2.3 Measures

5.2.3.1 Independent Variables

Undergraduate Grade Point Average (UGPA). UGPA was obtained through the university administrative system. In cases where UGPA was not registered, self-reported UGPA was used instead. Since the participants had UGPAs from different higher education systems, we brought them on one scale. We converted[20] their UGPAs on original scales from their countries into UGPAs on a US scale from 1 to 4.

Graduate Record Examination (GRE) General Test, Paper Version. The GRE General Test measures student abilities that are important at graduate level within any field of study (ETS, n.d.-c). The General Test consists of three sections: verbal reasoning (GRE-V), quantitative reasoning (GRE-Q), and analytical writing (GRE-AW). GRE-V assesses a wide range of abilities related to understanding, analyzing, and interpreting the kinds of texts commonly included in graduate schools' curricula. GRE-Q measures basic mathematical skills and ability to answer questions using quantitative methods,

20 In another study within this thesis, we used percentile ranks within each higher education system instead of converting the grades into one scale. In that study, we had enough participants to form percentile ranks for students from different countries (at least 20). However, the total sample size of 75 students does not allow us to form groups of students from specific countries. Conversion of grades into one scale is quite common in research on this topic (see the studies of Schwager et al., 2015 and Zimmermann et al., 2017).

and GRE-AW assesses critical thinking and analytical writing skills (ETS, n.d.-b). GRE scores are reported for GRE-V on a scale of 130–170, in 1-point increments, for GRE-Q, on a scale of 130–170, in 1-point increments, and for GRE-AW, on a scale of 0–6, in half-point increments.

5.2.3.2 Dependent variables

Graduate Grade Point Average (GGPA). GGPA is registered on the Dutch grading scale from 1 to 10 with 5.5 as a passing grade. For those students, who had not finished their master's yet, a proxy of GGPA was calculated based on components of the curriculum, which they had already finished. It was chosen to calculate proxies only for those students who completed at least 15 ECTS credits of their master's program, because this is the minimal number of credits obtained for the mandatory courses before starting internships. In our sample, only four students earned as little as 15 ECTS credits. The mean was 95 credits, and the median was 103.5 credits. Therefore, most of the students were close to obtaining the required 120 credits for their master's program. There were three students who did not earn any credits and three students who earned less than 15 credits. These six students were excluded from the analysis.

Internship Grade at the GSLS. This variable represents a grade for the performance during the research internship of nine months and 51 ECTS credits (and reporting on it). Research internships of such length are an obligatory requirement only at one graduate school (the GSLS) out of three participated graduate schools. These internships are typically performed in a laboratory with full immersion in research practice. During the internship, students implement a wide range of research tasks including the design, implementation, and reporting on a research project. The grade is given on the Dutch grading scale from 1 to 10 with 5.5 as a passing grade.

Rubrics at the GSLS. The GSLS applies three rubrics with the aim to facilitate the supervisors in giving comprehensive feedback to their masters' students on the quality of their performance during internships and to provide general criteria and standards for this feedback (Postmes et al., 2022). These rubrics are: (a) "research skills" with three main categories: "performing research", "practical skills", and "professional attitude". (b) "research report" with three main categories: "content", "structure and style", and "professional attitude"; (c) "presentation" with three main categories: "content", "presentation technique", and "composition and design". Each main category consists of several criteria.

The criteria of these three rubrics were developed following the learning objectives of the internship and refined during a number of pilots (Postmes et al., 2022). Notably, internship grades are not calculated from the supervisors' scores on rubrics. The rubrics are meant to provide justifications for the overall internship grade, but they constitute a separate assessment instrument. For this study, we chose to focus on those main categories of the rubrics, which are most closely related to research skills of students: "performing research", "practical skills" "content" (of a research report), and "structure

and style" (of a research report).

"Performing Research" Main Category. This category consisted of three criteria: "design research plan/experiments", "data analysis and interpretation", and "discussion research outcomes". Each criterion is assessed following the three levels quality descriptors: insufficient, satisfactory, and excellent. Following the approach of Postmes et al. (2022), we converted the supervisors' ratings on the rubric into five scores: 1.0 for insufficient level, 1.5 when both insufficient and sufficient levels were indicated, 2.0 for sufficient level, 2.5 for both sufficient and excellent, and 3.0 for only excellent level. Cronbach's alpha of "performing research" was acceptable: α = .79.

"Practical Skills" Main Category. This category had four criteria: "technical skills", "efficiency", "organization lab journal/log/work records", and "organization working space". The quality descriptors and their conversion into the metric scale were similar to what we described for the main category "performing research". Cronbach's alpha of "practical skills" was acceptable: α =.76.

"Content of Research Report" Main Category. The following criteria constitute this category: "title", "abstract", "Layman's summary", "introduction", "methods section", "results", "tables and figures", "discussion and conclusion". The quality descriptors and their conversion into the metric scale followed the principle described above. Cronbach's alpha of "content of research report" was good: α = .83.

"Structure and style" Main Category. This category consisted of three criteria: "structure and line of reasoning", "referencing", and "writing skills". The quality descriptors were the same as for other categories as well as their conversion into the metric scale. Cronbach's alphas of "structure and style" category was acceptable α = .77.

5.2.3.3 Control Variables

Socioeconomic Status (SES). Students were asked to indicate the level of the education of each of their parents. The highest level of parental education of either parent was derived as an indicator of SES (following Gooding, 2001; OECD, 2017). The scale was then transformed as closely as possible to International Standard Classification of Education (ISCED) and European Qualifications Framework (EQF), resulting into the following ordinal scale: 1= less than secondary education, 2 = upper secondary education, 3 = short-cycle tertiary education, 4 = bachelor's or equivalent level, 5 = master's or equivalent level, 6 = doctoral or equivalent level. We also gathered information on other indices of students' SES such as yearly family income, parental occupation, and self-perceived SES.

Number of Graduate ECTS Credits. To control for the fact that not all students earned the same amount of ECTS credits during their graduate program, we included number of ECTS credits as a control variable in the last step of the analysis. This allowed us to adjust for the possibility that students with lower number of ECTS credits after 26 months are those, who take longer time to earn higher GGPAs and vice versa.

5.2.4 A-priori Power Analysis

We conducted an a-priori power analysis with five predictors: UGPA, three scales of the GRE, and SES. We based the calculation on the effect sizes found in two studies in the European context (Schwager et al., 2015; Zimmermann et al., 2017). When predicting GGPA, the first study found an effect of .14 on a sample of 282 students, and the second study found an effect of .19 on a sample of 369 students. The power calculations are presented in Table 1. As follows from Table 1, to detect small to medium effect, our sample size must be above 154. After the announcements of this study, 234 students signed up to participate, but only 167 students appeared at the testing event.

Table 1 | *A Priori Power Analysis.*

Power	Effect size	Effect size (R^2)	Effect size (f^2)	Sample size
0.95	Medium	.13	.15	139
0.80	Medium	.13	.15	92
0.95	Small–medium	.08	.09	234
0.80	Small–medium	.08	.09	154
0.95	Small	.02	.02	975
0.80	Small	.02	.02	635

5.2.5 Participants

We collected data on 167 students from three graduate schools at the start and at the end of their master's program. After the exclusion of six students who earned less than 15 ECTS credits within two and a half years of being enrolled in their master's program, we formed a sample of 161 students (Sample 1). On this sample, we tested the predictive validity of the GRE above UGPA toward one dimension of graduate study success, namely, GGPA. For the analysis of other dimensions of graduate study success (internship grade and rubrics' main categories), we formed a subsample of the original sample (Sample 2), which consisted of 75 GSLS students who completed their nine-months internship, meaning that their grades and assessments on rubrics were available as well. The characteristics of each sample are provided in Table 2.

5.2.6 Data Analysis

The regression analyses were conducted in SPSS version 26. Missing values were handled by using the expectation maximization (EM) technique, because the percentage of missingness was no larger than 5%. First, the hierarchical regression analyses were used. We had six outcome variables. The model with the first outcome variable (GGPA) was analyzed, using Sample 1. The models with the other five outcome variables (Internship grade and four categories of rubrics) were analyzed using Sample 2[21]. In the first step of

21 Even though we are aware of the possible issue of multiple testing that might arise from running five separate regressions on Sample 2, we decided not to control for it at this stage of research, because Sample 2

each regression, we included UGPA. In the second step, we added three sections of the GRE and in the third step, we controlled for SES. The analysis with GGPA as an outcome variable had an additional fourth step, in which we controlled for number of graduate ECTS credits.

Table 2 | *Samples' Demographical, Socioeconomic, and Educational Characteristics.*

Characteristics	**Sample 1 (N = 161)**	**Sample 2 (N = 75)**
Gender: females	53%	69%
Age	*M* = 23.0 *SD* = 2.4 Median = 22.4	*M* = 22.9 *SD* = 2.0 Median = 22.0
Graduate school		
GSLS	n = 93	n = 75
GSNS	n = 54	0
GSGS	n = 14	0
Country of origin		
Netherlands	66%	59%
Other EU	19%	29%
Outside of EU	14%	12%
Students with at least one parent with a higher education degree	69%	69%
Parental income		
No knowledge of it	36%	43%
Less than 25,000€	11%	9%
25,000€ – 50,000€	16%	15%
50,000€ – 75,000€	14%	8%
75,000€ – 100,000€	8%	9%
100,000€ – 150,000€	9%	13%
More than 150,000€	6%	3%
Self-perceived SES on a scale from 1 (low) to 10 (high)	*M* = 6.7 *SD* = 1.5 Median = 7.0 Range = 3–10	*M* = 6.8 *SD* = 1.4 Median = 7.0 Range = 3–10
Previous HEI		
The same research university they enrolled for their masters'	45%	41%
Another research university in the Netherlands	18%	23%
University of applied sciences in the Netherlands	5%	1%
HEI outside of the Netherlands	24%	27%
Other type of HEI	8%	8%
Missingness (values)	0.2%	4.8%

We set alpha level of .10 and considered it as marginal significance, because according to our a-priori power analysis (Table 1), the sample size of Sample 1 (n = 167) is not big enough to detect small effect sizes under α-level 0.05 and the sample size of Sample 2 (n = 75) is not big enough to detect small and small to medium effect sizes under α-level 0.05. In other words, we were willing to increase Type I error (we take a slightly higher risk to detect an effect, which is not there) in order to decrease Type II error (we take a lower

has only 75 cases and our study is exploratory at this stage. As soon as we get more data from the rest of the students, who have not graduated yet, we might consider to statistically account for multiple testing.

risk of missing an effect, which is actually there). We made this decision because of the exploratory nature of this study toward the dimensions of graduate study success, related to quality of independent research work of graduate students.

We supplemented the regression analysis by presenting the figures (Figures 2.1–Figure 7.2), two per each outcome (i.e., per dimension of graduate study success). The first type of figures (with an extension .1 at the end, namely Figures 2.1, 3.1, 4.1, 5.1, 6.1, and 7.1) allowed us to contrast the predicted score of each individual on outcome (X-axis) against actual score of this individual on this outcome (Y-axis). On the left side of each figure, the predictions (X-axis) were based only on UGPA. On the right side of each figure, the predictions were based on UGPA *and* three GRE scales. If the predictions had completely corresponded to the actual values, we could expect all data points have the same score on X-axis as on Y-axis.

The second type of figures (with an extension .2 at the end, namely Figures 2.2, 3.2, 4.2, 5.2, 6.2, and 7.2) was designed to show histograms of the absolute values of residuals (differences between predicted and actual scores). Similar to the first type of figures, the left side of a figure (a model with only UGPA as a predictor) was designed for comparison with its right side (a model with both UGPA and GRE). We suggest to the reader to interpret the figures as follows: The better prediction model has more cases with residuals being close to zero, the worse prediction model has less cases with residuals being close to zero. Overall, both types of figures visualized what predictive validity of GRE means for each of the examined outcomes. The graphs were designed in Stata.

5.3 Results

5.3.1 Predicting GGPA

The analyses presented below are conducted on Sample 1: the full sample of students from three STEM graduate schools. Table 3 shows an overview of bivariate correlations between the study variables of Sample 1. The scores on all GRE sections are positively intercorrelated. Also, GRE-V and GRE-AW (but not GRE-Q) are correlated with UGPA. Three sections of the GRE, UGPA, and ECTS credits achieved during graduate studies are all positively related to GGPA. We also note that SES is positively related to GRE-V and, on a marginal level, to GGPA.

Table 3 | *Intercorrelations between Study Variables of Sample 1.*

Variable	*n*	*M*	*SD*	1	2	3	4	5	6	7
1. GRE-V	161	151.83	6.27	1						
2. GRE-Q	161	153.91	6.86	.44***	1					
3. GRE-AW	161	3.57	0.57	.43***	.28***	1				
4. UGPA	159	3.33	0.50	.31***	.13	.33***	1			
5. GGPA	161	7.76	0.57	.45***	.36***	.39***	.42***	1		
6. SES	161	4.15	1.32	.16*	.10	-0.01	0.08	.15†	1	
7. Graduate ECTS credits	161	96.86	31.25	-.17*	.05	.02	-.03	.17*	.06	1

Note. †$p < .10$, *$p < .05$. **$p < .01$, *** $p < .001$.

5.3.1.1 Regression Analysis of Sample 1

The results for incremental validity of the GRE scores are presented in Table 4. The GRE sections predict beyond and above of what UGPA predicts in GGPA: Each of three GRE sections is positively related to GGPA (see Step 2). Namely, with each additional score on GRE-V, GRE-Q, and GRE-AW, GGPA increases with 0.02 (B = 0.02, SE = 0.01, t = 2.63, $p = .010$), 0.02 (B = 0.02, SE = 0.01, t = 2.60, $p = .010$), 0.15 (B = 0.15, SE = 0.07, t = 2.03, $p = .044$) respectively. Altogether, three GRE sections add 16% of explained variance in GGPA to what has already been explained by UGPA.

To illustrate the findings, let us examine the figures. Figure 2.1 visualizes the findings that data points are predicted better when both UGPA and GRE are included in the model. Figure 2.2 illustrates that once GRE is included together with UGPA into the model (the right side of the figure), more individual cases have smaller residuals.

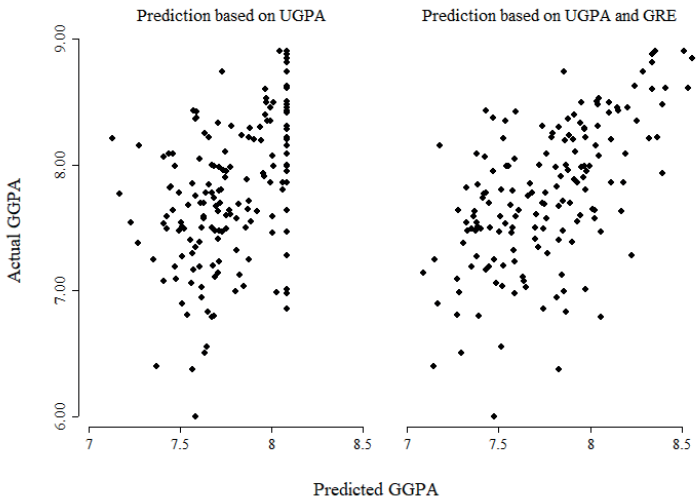


Figure 2.1 | *Predicted Individual Scores versus Actual Scores of GGPA.*

Table 4 | *Hierarchical Multiple Regression Analyses Predicting GGPA.*

Predictor	GGPA	
	ΔR²	**β**
Step 1	.18***	
UGPA		.42***
Step 2	.16***	
UGPA		.28***
GRE-V		.21*
GRE-Q		.19*
GRE-AW		.15*
Step 3	.01	
UGPA		.28***
GRE-V		.20*
GRE-Q		.19*
GRE-AW		.16*
SES		.08
Step 4	.04**	
UGPA		.28***
GRE-V		.26**
GRE-Q		.16*
GRE-AW		.14†
SES		.06
Graduate ECTS credits		.21**
Total R^2	.39	

Note. † $p < .10$, *$p < .05$. **$p < .01$, **1* $p < .001$. n = 161. UGPA – Undergraduate Grade Point Average; GRE-V – Graduate Record Examinations Verbal Reasoning; GRE -Q - Graduate Record Examinations Quantitative Reasoning; GRE -AW – Graduate Record Examinations Analytical Writing; SES – socio-economic status, operationalized as the highest level of parental educational achievement.

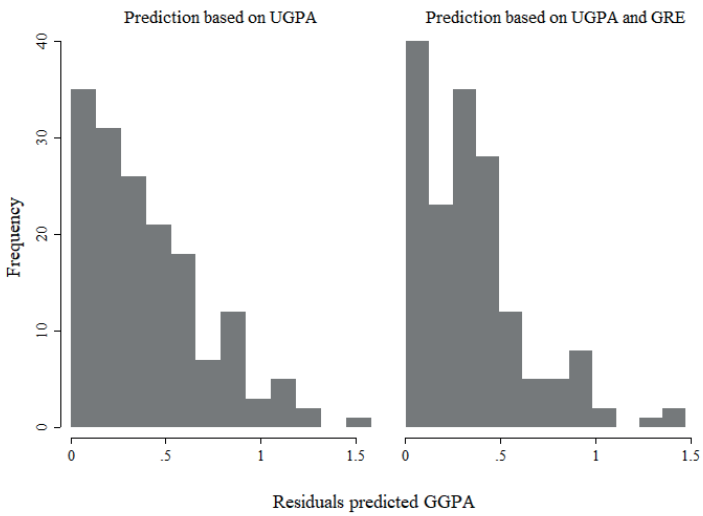


Figure 2.2 | *The Absolute Value of Difference between Predicted and Observed Scores of GGPA.*

Finally, we find that even after the addition of SES, the detected positive relationships between GRE sections and GGPA hold, and this addition of SES explains negligible amount of variance (Table 4, Step 3). The detected relationships hold also after addition of ECTS credits, achieved during graduate program; however, the relationship between GRE-A and GGPA attenuates with inclusion of ECTS credits into the model, GRE-A becomes marginally significant in the last model (Table 4, Step 4).

5.3.2 Predicting Five Internship-Related Dimensions of Graduate Study Success

The analyses that we present below are based on 75 GSLS students (Sample 2), all of whom had completed an internship of 51 ECTS credits during nine months. In addition to the grade, they received supervisors' ratings on their internship performance within rubrics. Table 5 shows the bivariate correlations between the study variables of Sample 2. The three GRE sections are significantly intercorrelated, all correlations are positive. GRE-V and GRE-AW (but not GRE-Q) have significant positive correlations with UGPA. GRE-V has significant positive correlations with all five examined dimensions of graduate study success, while GRE-Q and GRE-AW are correlated with four examined dimensions of graduate study success, except the "content" rubric. SES (defined as the highest education level of either parent) has a significant correlation with one study variable, namely, the rubric "performing research".

Table 5 | *Intercorrelations between Study Variables of Sample 2.*

Variable	*n*	*M*	*SD*	1	2	3	4	5	6	7	8	9	10
1. GRE-V	75	150.83	6.03	1									
2. GRE-Q	75	151.89	6.35	.52***	1								
3. GRE-AW	75	3.61	0.61	.51***	.36**	1							
4. UGPA	75	3.42	0.50	.35**	.16	.23*	1						
5. Internship grade	75	7.86	0.70	.43***	.43***	.35**	.28*	1					
6. Rubric "Performing research"	67	2.12	0.37	.43***	.22†	.21†	.24*	.74***	1				
7. Rubric "Practical skills"	66	2.32	0.30	.26*	.31*	.27*	.16	.65***	.59***	1			
8. Rubric "Content"	66	2.17	0.27	.24†	.10	.11	.24†	.56***	.57***	.59***	1		
9. Rubric "Structure and Style"	65	2.18	0.36	.22†	.29*	.22†	.23†	.51***	.34**	.41**	.72***	1	
10. SES	75	4.31	1.34	.07	-.04	.02	.10	.14	.22†	.08	.16	.01	1

Note. †$p < .10$, *$p < .05$. **$p < .01$, *** $p < .001$.

5.3.2.1 Regression Analysis of Sample 2

Table 6 shows the gain in explained variance after the addition of three GRE sections into the models with UGPA. Each model is related to one dimension of graduate study success. Table 6 also provides regression coefficients of each predictor, included into the model. As the last step in this table, the results for the models after the addition of SES are presented.

Predicting Internship Grade. The GRE sections predict above and beyond what UGPA predicts in internship grade: They add 20% of explained variance in internship grade to what has already been explained by UGPA. GRE-Q was found to be a significant positive predictor: As GRE-Q increases with one score, our model predicts that internship grade will increase with 0.03 score (B = 0.03, SE = 0.01, t = 2.29, p = .025). GRE-V and GRE-AW were not significant predictors. Figure 3.1 and Figure 3.2 illustrate the improvements, to which the addition of GRE scales lead: More cases are better predicted once GRE scales are in the model. The addition of SES does not chance this relationship.

Table 6 | *Hierarchical Multiple Regression Analyses Predicting Internship Grade and Scales of Internship Rubrics.*

Predictor	Internship Grade		Scale "Performing Research"		Scale "Practical Skills"		Scale "Content"		Scale "Structure and Style"	
	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β
Step 1	.08*		.05†	.22†	.03		.04†		.04†	
UGPA		.28*				.17		.21		.20
Step 2	.20***		.15**		.12*				.09†	
UGPA		.15		.08		.08	.03	.14		.14
GRE-V		.18		.42**		.06		.21		-.03
GRE-Q		.27*		-.02		.22†		-.02		.26†
GRE-AW		.12		-.02		.17		-.02		.13
Step 3	.01		.03		<0.01		.01		<.01	
UGPA		.14		.06		.07		.13		.14
GRE-V		.17		.41**		.05		.20		-.03
GRE-Q		.29*		<.01		.23†		-.01		.26†
GRE-AW		.13		-.01		.17		-.02		.13
SES		.12		.16		.04		.10		<0.01
Total R^2	.29		.22		.15		.08		.13	

Note. † p < .10, *p < .05. **p < .01, *** p < .001. n = 75. UGPA—Undergraduate Grade Point Average; GRE-V—Graduate Record Examinations Verbal Reasoning; GRE-Q—Graduate Record Examinations Quantitative Reasoning; GRE-AW— Graduate Record Examinations Analytical Writing; SES — socio-economic status, operationalized as the highest level of parental educational achievement.

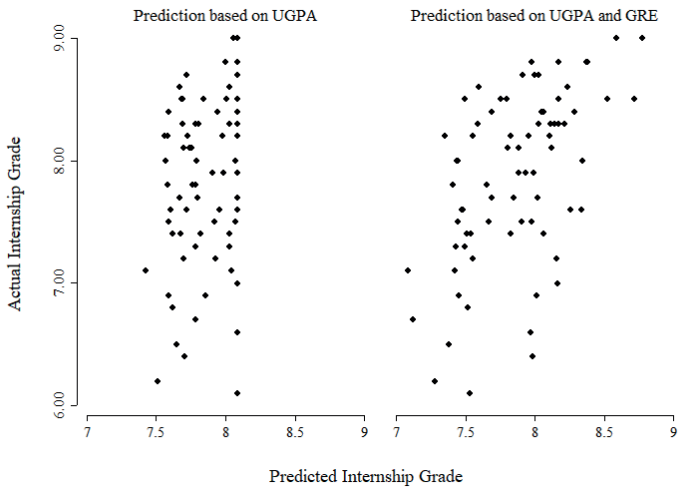


Figure 3.1 | *Predicted Individual Scores versus Actual Scores of Internship Grade.*

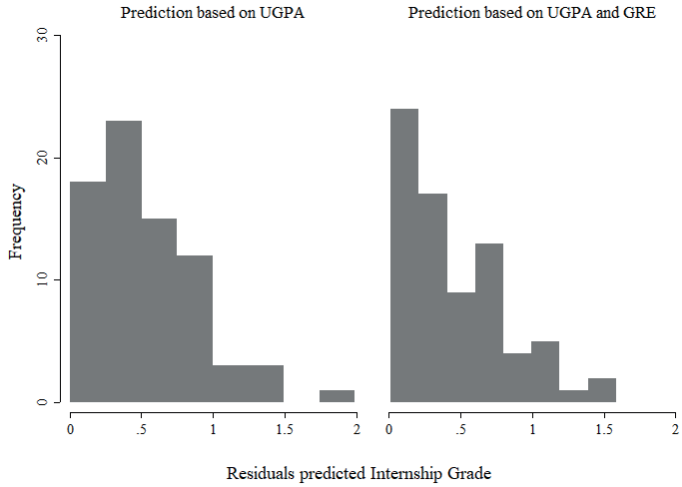


Figure 3.2 | *The Absolute Value of Difference Between Predicted and Observed Scores of Internship Grade.*

Predicting the Supervisors' Scores on "Performing Research" Category. The addition of three GRE sections to UGPA in the model significantly improves the predictive power of the model: It adds 15% to explained variance in the supervisors' scores on "performing research". GRE-V was found as a significant positive predictor: With each additional score on GRE-V, the supervisors' ratings on "performing research" category increased with 0.03 (B = 0.03, SE = 0.01, t = 2.98, p = .004). Two other GRE scales were not significant. Figure 4.1 and Figure 4.2 visualize the better predictability once both UGPA and GRE are in the model. The addition of SES to the model did not change the detected relationships.

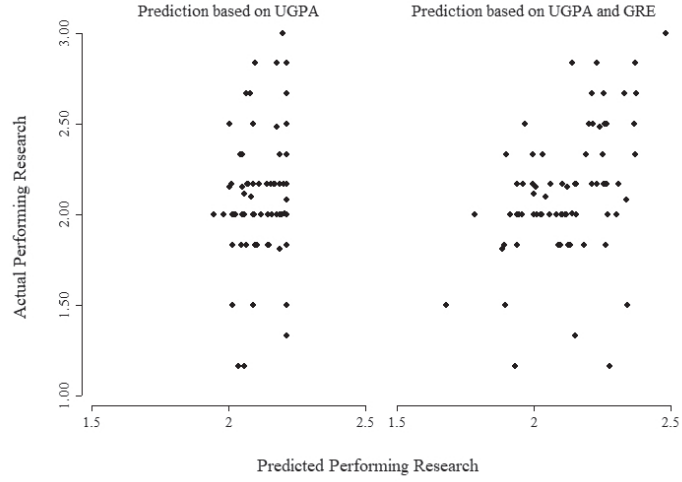


Figure 4.1 | *Predicted Individual Scores versus Actual Scores of "Performing Research" Category of "Research Skills" Rubric.*

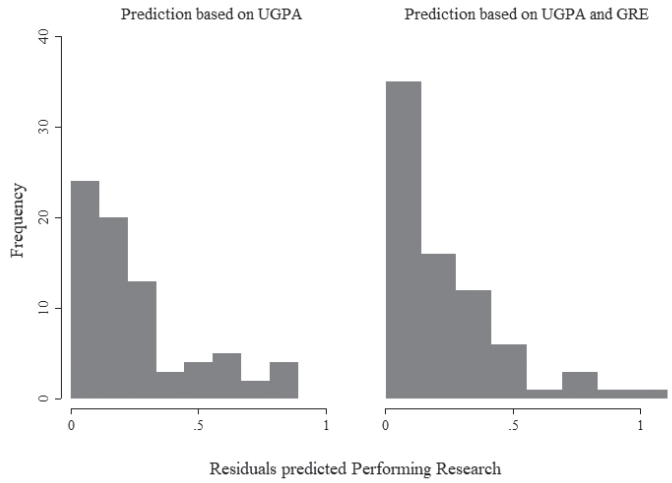


Figure 4.2 | *The Absolute Value of Difference between Predicted and Observed Scores of "Performing Research" Category of "Research Skills" Rubric.*

Predicting the Supervisors' Scores on "Practical Skills" Category. The addition of the three GRE sections significantly improves the prediction of supervisors' ratings on the "practical skills" category. The strongest predictor out of the three GRE scales appears to be GRE-Q, though on a marginal level: With each score on GRE-Q, the supervisors' ratings on "practical skills" category increased with .01 score (B = 0.01, SE = 0.01, t = 1.70, p = .093). All other predictors were nonsignificant. Figure 5.1 and Figure 5.2 show that once GRE is in the model, more cases are predicted closer to the actual values. These findings do not change after accounting for SES.

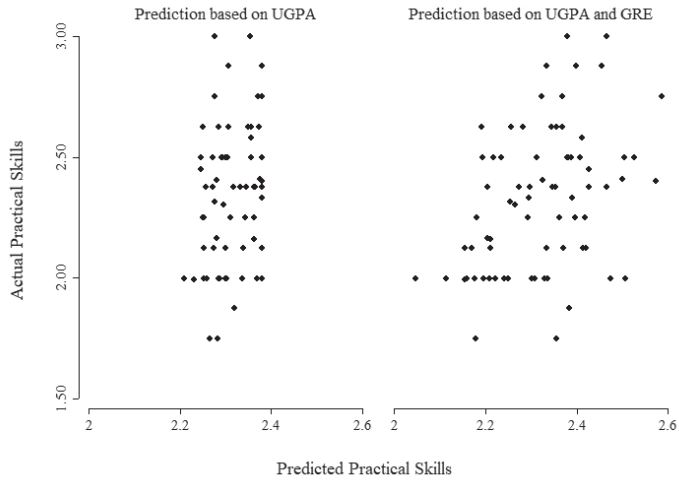


Figure 5.1 | *Predicted Individual Scores versus Actual Scores of "Practical Skills" category of "Research Skills" Rubric.*

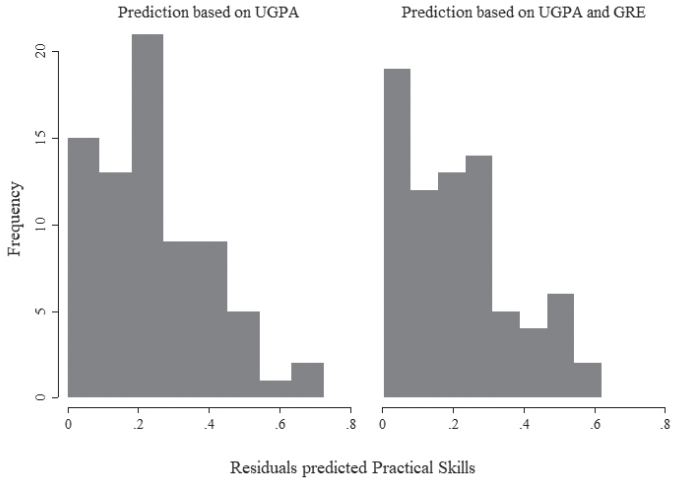


Figure 5.2 | *The Absolute Value of Difference between Predicted and Observed Scores of "Practical Skills" Category of "Research Skills" Rubric.*

Predicting the Supervisors' Scores on "Content" Category. The GRE did not add a significant amount of explained variance to what has already been explained by UGPA. Also, none of the regression coefficients for the GRE reached the significance level (all $p > .100$). Visual inspection of Figure 6.1 and Figure 6.2 is in line with these findings. Though some changes are observed between the left and right sides of the figures, they are minor, which does not allow us to interpret them as an improvement once GRE is added.

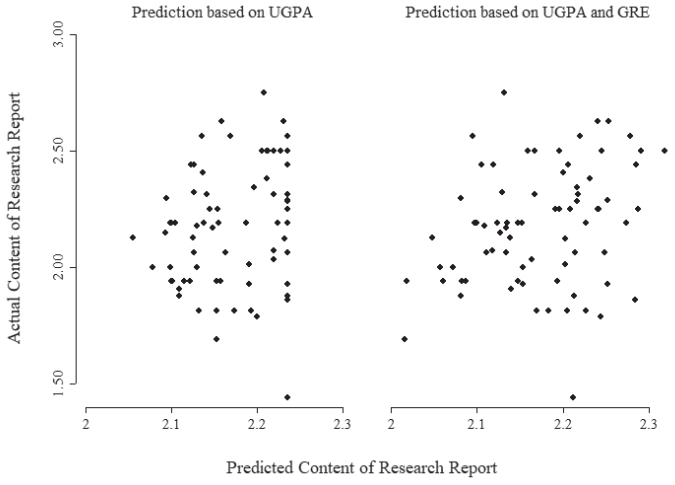


Figure 6.1 | *Predicted Individual Scores versus Actual Scores of "Content" Category of "Research Report" Rubric.*

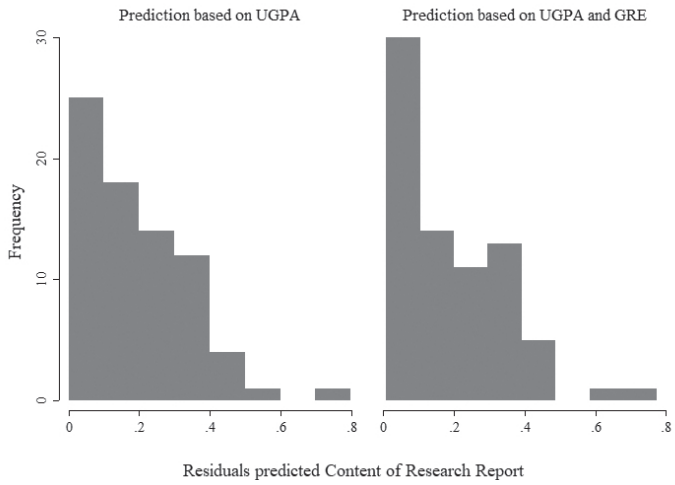


Figure 6.2 | *The Absolute Value of Difference between Predicted and Observed Scores of "Content" Category of "Research Report" Rubric.*

Predicting the Supervisors' Scores on "Structure and Style" Category. The three sections of the GRE marginally improved the model with only UGPA (the improvement of 9% in explained variance). GRE-Q was found as a marginally significant predictor: With each additional score on GRE-Q, supervisors' ratings on "structure and style" category increased with .01 (B = 0.01, SE = 0.01, t = 1.98, p = .052). All other predictors were nonsignificant. Figure 7.1 and Figure 7.2. visualize the improvements achieved, once GRE is in the model. The addition of SES to the model does not alter the detected marginally significant relationship.

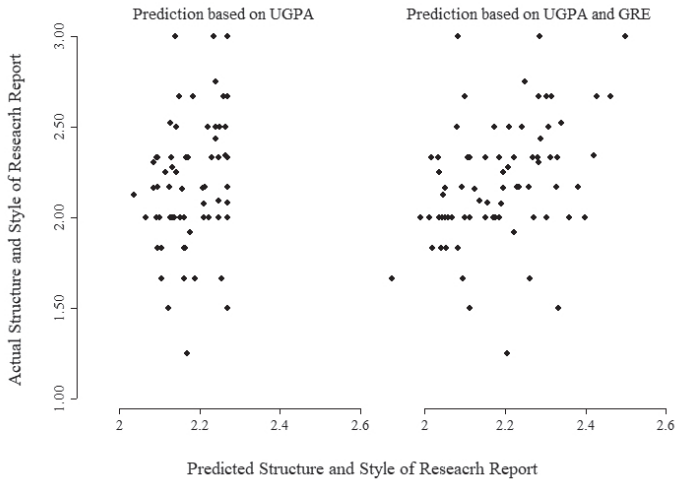


Figure 7.1 | *Predicted Individual Scores versus Actual Scores of "Structure and Style" Category of "Research Report" Rubric.*

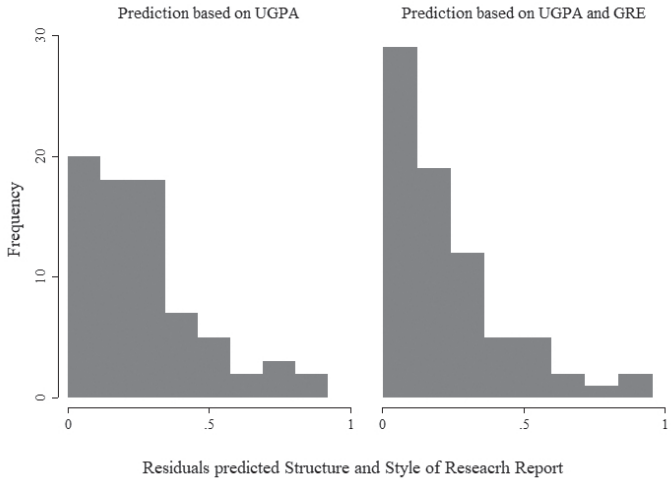


Figure 7.2 | *The Absolute Value of Difference between Predicted and Observed Scores of "Structure and Style" Category of "Research Report" Rubric.*

5.4 Discussion

We examined the incremental validity of the GRE beyond and above UGPA in research masters' programs and whether the detected relationships hold after accounting for SES. Overall, this study showed that the GRE is a meaningful predictor of graduate study success for STEM research masters' programs, and the detected relationships held after accounting for SES. Importantly, this study was conducted on a sample of students, who were not selected based on GRE scores, and thus the range of the GRE scores in our sample was wide and comparable to the range of future potential applicants in HEIs without a pre-set threshold. Moreover, the participants were students enrolled in more than 20 research masters' programs in STEM-related fields at three large graduate schools, which allows the generalizability of our findings to a wide range of STEM masters' programs.

5.4.1 GRE as a Predictor of GGPA. The Role of SES in This Relationship

Our result that the GRE is predictive of GGPA beyond and above UGPA replicated a widespread finding on a positive relation of the GRE toward GGPA in STEM disciplines (Klieger et al., 2014; Moneta-Koehler et al., 2017; Zimmermann et al., 2017), confirmed also by meta-analytical findings (Kuncel et al., 2010). We also showed that even though SES is correlated to GRE-V and marginally to GGPA, the relationship between three GRE sections and GGPA did not significantly drop once SES was accounted for.

To place this finding in the context of the existing literature on the topic, we followed the approach of Schwager et al. (2015) and computed partial correlations. Namely, we

compared partial correlations between the GRE sections and GGPA after controlling for SES, to the zero-order (Pearson) correlations, reported in Table 4. What we found is that after controlling for SES, the correlations between GRE and GGPA stayed effectively the same: partial $r_{GRE\text{-}V\text{-}GGPA}$= .43, p < .001 instead of Pearson correlation r= $r_{GRE\text{-}V\text{-}GGPA}$= .45, p < .001; partial $r_{GRE\text{-}Q\text{-}GGPA}$= .35, p < .001 instead of a Pearson correlation r= $r_{GRE\text{-}Q\text{-}GGPA}$= .36 p < .001; partial $r_{GRE\text{-}AW\text{-}GGPA}$= .39, p < .001 identical to a Pearson correlation $r_{GRE\text{-}AW\text{-}GGPA}$= .39 p < .001. These findings are in line with the primary study of Schwager et al. (2015) on students from taught masters' programs at a Dutch university as well as with the meta-analysis of Sackett et al. (2009) on the relationship between standardized undergraduate test (Scholastic Aptitude Test; SAT) and UGPA. In both studies, the reductions in correlations between scores on standardized tests and GPA after controlling for SES ranging from negligible to small. Hence, our findings replicate the findings both on undergraduate and graduate level.

5.4.2 GRE as a Predictor of Dimensions of Graduate Study Success, Related to Independent Research. The Role of SES in These Relationships

In addition to examining the grades for research work (i.e., internship grade), we also examined the rubric categories as dimensions of graduate study success. This allowed us to shed light on whether the GRE has any relation to various quality aspects of students' research work. We found evidence that the GRE had incremental predictive value above and beyond UGPA in predicting not only internship grades, but also supervisors' ratings on three rubric categories: "performing research", "practical skills" (marginally), and "structure and style of research report" (marginally). The GRE did not have incremental predictive value toward supervisors' ratings on rubric category "content of research report".

To the best of our knowledge, this is the first study to examine these questions and therefore, we do not have prior literature to relate to our findings. However, Zimmermann et al. (2015) looked at the relationship between the GRE scales and thesis grades on masters' programs at a Swiss university and did not find one. The authors of that study see the possibility that "the grade earned for the thesis is strongly influenced by variations in the grading-schemes used among academic supervisors. That is, the grades assigned by different supervisors might stand for quite different achievements and, thus, are rather not comparable" (Zimmermann et al., 2015, p. 19). This is an example of an attenuating factor, when unreliability of a study success measure reduces the estimate for a relationship between standardized tests and study success (Kuncel & Hezllet, 2007). In our study, the rubric categories, consisting of items on specific aspects of students' research work, were used by all supervisors to provide justification for internship grades. We assume that this contributed to acceptable reliabilities in the use of these rubrics' categories and to the fact that the relationships between individual GRE scales and research-related dimensions of graduate study success were detected.

Importantly, not all GRE scales were linked to research-related dimensions of graduate study success. GRE-Q was the scale which significantly related to internship grade and supervisors' ratings on "practical skills" and "structure and style of research report". GRE-V was the scale which significantly related to supervisors' ratings on "performing research". GRE-AW did not predict any of the research-related dimensions of graduate study success. It could be that our sample size of 75 students was not big enough, and we missed the possible significant relationships of small or small to medium size. It could also be that only one or two and not all three GRE scales are related to the examined dimensions. Future research could explore why not all three scales are indicative of research-related dimensions of graduate study success.

We also found that these described relationships between GRE scales and research-related dimensions of graduate study success hold after accounting for SES. Like with GGPA, we computed partial correlations between three GRE sections and five examined research-related dimensions of graduate study success, controlled for SES. We compared these partial correlations to Pearson uncorrected zero-order correlations and revealed that they were practically the same[22]. This, together with the negligible amount of added explained variance after inclusion of SES in our models, indicated that the detected relationships between the GRE sections and internship grade and supervisors' ratings on "performing research", "practical skills", and "content and structure of research report" were not artifacts of SES.

5.4.3 Strengths of the Study and Future Directions of Research

The first strength of this study is the variability in the GRE scores: On each GRE scale we had a wide variety of scores, especially on GRE-V (from 133 to 167) and GRE-Q (from 134 to 170), which corresponds almost to the whole range of possible scores on these scales. This is in contrast to many studies in this field, which were often conducted on students, already selected for admissions based on the GRE, and this restricted range in the scores on the predictor might have affected their findings (e.g., Cox et al., 2009; Sandford, 2009). The wide range of scores in GRE allows us to transfer our findings more easily to the population of our interest: the diverse pool of applicants to research masters' programs, who will naturally have a potentially wide range of GRE scores, provided the institution does not set a minimum threshold score for application.

The second strength is that this is the first study to explore the predictive power of the GRE on quality-related aspects of independent research work of graduate students. From our study, it appears that individual scales of the GRE are predictive of several aspects of graduate student research work. Future research could consider exploring why the GRE predicts specific research skills (e.g., which cognitive constructs, measured by the GRE, affect the level of research skills?).

22 Available upon request

Finally, this study is only the third to be conducted on the GRE in the European context and extends our understanding of how the GRE can be applied in STEM graduate schools in Europe, considering the distinct position of master's (as a final degree) within the European three cycle system (Schwager et al., 2015; Zimmermann et al., 2017). The possibilities and conditions of this application will be discussed in more detail in the section "Practical Implications for Graduate Admissions in Europe".

5.4.4 Limitations

This study does not come without limitations. The first and, perhaps, the main limitation is that it was conducted in a low-stake situation: The students had been already admitted to their master's program, so their performance on the GRE did not affect their admissions on a master's program and they did not prepare for the test as they would usually do in a high-stake situation. According to students' self-reports, the median preparation time was 20 minutes, which mostly included reading information on the GRE test content and structure before taking it. On the one hand, the fact that the participants did not prepare for the test, placed them in equal positions and provided realistic assessment of their verbal, quantitative, and analytical writing skills. On the other hand, if students' admissions to their desired masters' programs would indeed depend on the GRE scores, we expect that they would report much higher preparation rates. SES could then have played a more significant role (with more wealthy students likely making more financial investments in buying courses and books for GRE preparation, hiring tutors, and being able to make more exam attempts to obtain the desired minimum level of scores). The debate on this issue would benefit from future studies on the GRE within a high-stake situation, conducted within the European context.

The second limitation is that we used only one aspect (namely, parental education) as a proxy for SES. We acknowledge that SES is a complicated construct (Rodríguez-Hernández et al., 2020), which goes beyond parental education only. However, parental education has been shown to represent SES adequately due to the greatest load it exerts on the SES index (Caro & Cortés, 2012) and is used in research similar to ours (e.g., Schwager et al., 2015). We initially also gathered information on other aspects of SES (parental occupational status, family financial situation, and subjective assessments of SES). Having tested those operationalizations as proxies for SES did not deliver substantially different results.

The third limitation is also related to SES. It might be that we found the relationship between the GRE scores and graduate study success to be independent from SES because we had a relatively homogeneous sample in terms of SES. Indeed, 69% of students in our sample had at least one parent with a higher education degree which is higher than the average of tertiary attainment in the Netherlands (38%) and the OECD average (39%; OECD, 2019). This perhaps indicates that SES does not play a major role in such a preselected sample, while it could have played a role in a more diverse sample.

5.4.5 Practical Implications for Graduate Admissions in Europe

Even though the results of this study demonstrate that the GRE appears to predict several dimensions of study success in STEM research masters' programs, several aspects should be considered when deciding whether to use GRE as a selection method. Some of these considerations speak for the usage of the GRE and the others speak against it. We suggest that each STEM graduate school weights these considerations carefully in the context of its own admissions goals and educational mission.

The considerations that speak in favor of implementation of the GRE are as follows. First, the GRE appears to predict not only traditional dimensions of study success (such as GGPA), but also shows a potential to predict the quality of graduate students' research performance. Second, the GRE places students from different educational backgrounds on one scale and therefore allows to compare their skills, required at the entrance to a graduate school. This fact not only facilitates the work of admissions committees in terms of their invested time and efforts, but also makes the admissions decisions transparent to different stakeholders (including applicants and their parents). Third, being able to submit the GRE scores might benefit students from not very well known and established HEIs across the world and within their own country (e.g., universities of applied sciences in the Netherlands). These students can get a serious consideration for research masters' programs in Europe, as their levels of verbal, quantitative, and analytical writing skills are then better understood by the admissions committees, who could have previously disregarded their application due to unfamiliarity with the applicant's HEI, the quality of its undergraduate programs and examinations, or by using low-validity and more subjective methods such as traditional interviews. Fourth, using the GRE provides a possibility for benchmarking: The HEIs can compare the level of skills between own applicants who graduated from their own bachelor programs and applicants from other countries and universities.

The considerations that speak against the usage of the GRE are as follows. First, as soon as the GRE is assigned a significant weight in admissions decisions, the scores might become prone to coaching effects, because students start preparing for the test. There is a study in the US context, however, which finds that mostly GRE-A is prone to coaching, but not the other two sections (Powers, 2017). More studies (also in the European context) are required to ensure that the effect of coaching is indeed limited. Second, the standardized tests are prone to *stereotype threat*—a self-fulfilling prophecy of test takers from minority groups that their underperformance on standardized tests might confirm the stereotypes of their minority group's intellectual capacity, which leads to actual underperformance (Garces, 2014; Posselt, 2016). Further research in the European context on this topic would be insightful for a better understanding of the extent to which stereotype threat affects the performance of European minority groups. The third consideration is that while the GRE is not established as a selection method in the European research-intense graduate schools, it might take additional time and effort to ensure its acceptability by stakeholders (e.g., applicants, admissions committees).

If the GRE is used, admissions committees should familiarize themselves with the best practices of implementing it in admissions processes. The ETS guidelines state that the GRE scores cannot be used as a sole or primary admissions method. Moreover, the ETS does not recommend applying cut-off scores and encourages admissions committees to account for measurement error using special tables (ETS, 2018). Posselt (2016) in her qualitative study showed how easily these recommendations are disregarded in admissions practices of the top graduate schools in the US.

5.4.6 Legal Limitations

Another important consideration for the GRE usage is legal requirements. The Dutch law, for example, does not allow HEIs to pass costs for (admissions) examinations on students with diplomas from EEA, equivalent to the Dutch diploma (Ministry of Education, Culture and Science of the Netherlands, 2021). Additionally, taking the GRE requires financial investments (standard GRE General test administration costs are $205.00 in 2021 in all areas of the world except China and India, where the prices are higher). According to Dutch law, passing the costs of admissions examinations on to an applicant is allowed only if (s)he does not meet the requirements of a higher educational program or does not hold a diploma that gives the right to admission (Ministry of Education, Culture and Science of the Netherlands, 2021). In practice, adhering to these requirements would mean that the GRE could have been either introduced optionally or required only from international students outside of European Economic Area (EEA). The optional usage of GRE might have a downside that only students with higher scores submit theirs (Posselt, 2016). Requiring GRE scores only from students with non-EEA diplomas will assist in comparing their skills between themselves but will not help in comparing their skills with those of students with bachelors' diplomas from within the EEA.

5.5 Conclusion

The results of our study indicate that GRE scores are predictive for graduate study success in English-taught research masters' programs at a major European university. All GRE scales appear to be predictive for GGPA. GRE-Q is predictive for internship grade and for supervisors' assessments of students' practical research skills and the structure and style of their research reports. GRE-V is predictive of supervisors' assessments of students' skills on performing research. The implementation of the GRE in admissions to selective European masters' programs, however, requires serious consideration of potential pros and cons as well as accounting for the limitations, imposed by the law.

5.6 References

Altman, R., Conrad, L., Flaugher, R., Thompson, R., Wallmark, M., & Willingham, W. (1977). In Conrad L., Trismen D., & Miller R. (Eds.), *Graduate Record Examinations: Technical manual* (pp. 1–104). Educational Testing Service.

European Commission, Education, Audiovisual and Culture Executive Agency, & Eurydice (2020). *The European Higher Education Area in 2020: Bologna Process Implementation Report*. Luxembourg: Publications Office of the European Union. https://eacea.ec.europa.eu/national-policies/eurydice/content/european-higher-education-area-2020-bologna-process-implementation-report_en

Bologna Working Group on Qualifications Frameworks. (2005). *A framework for qualifications of the European Higher Education Area.* Ministry of Science, Technology and Innovation of Denmark. Copenhagen. https://edudoc.ch/record/46590?ln=en

Bridgeman, B., Burton, N., & Cline, F. (2009). A note on presenting what predictive validity numbers mean. *Applied Measurement in Education*, *22*(2), 109–119. https://doi.org/10.1080/08957340902754577

Burmeister, J., McSpadden, E., Rakowski, J., Nalichowski, A., Yudelev, M., & Snyder, M. (2014). Correlation of admissions statistics to graduate student success in medical physics. *Journal of Applied Clinical Medical Physics*, *15*(1), 375–385. https://doi.org/10.1120/jacmp.v15i1.4451

Burton, N. W., & Wang, M. (2005). *Predicting long-term success in graduate school: A collaborative validity study.* (Report No. 99-14R. ETS RR-05-03). Educational Testing Service. http://grad.uga.edu/wp-content/uploads/2017/09/GRE_Research_Report.pdf

Burton, N., & Wang, M. (2014). Predicting long-term success in graduate school: A collaborative validity study. In C. Wendler & B. Bridgeman (Eds.), *The Research Foundation for the GRE revised General Test: A Compendium of Studies* (pp. 5.6.1 –5.6.5). Educational Testing Service. https://www.ets.org/s/research/pdf/gre_compendium.pdf

Caro, D. H., & Cortés, D. (2012). Measuring family socioeconomic status: An illustration using data from PIRLS 2006. In *Issues and Methodologies in Large-Scale Assessments* (Vol. 5, pp. 1–25). Educational Testing Service and International Association for the Evaluation of Educational Achievement. https://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_05_Chapter_1.pdf

Caspersen, J., Frolish, N., Karlsen, H., & Aamodt, P. O. (2014). Learning outcomes across disciplines and professions: Measurement and interpretation. *Quality in Higher Education*, *20*(2), 195–215. https://doi.org/10.1080/13538322.2014.904587

Clayton, V. (2016). *The Problem with the GRE*. https://www.theatlantic.com/education/archive/2016/03/the-problem-with-the-gre/471633/

Cox, G. W., Hughes Jr., W. E., Etzkorn, L. H., & Weisskopf, M. E. (2009). Predicting computer science Ph.D. completion: A case study. *IEEE Transactions on Education*, *52*(1), 137–143. https://doi.org/10.1109/TE.2008.921458

Croizet J.-C. & Dutrevis M. (2004). Socioeconomic status and intelligence: Why test scores do not equal merit. *Journal of Poverty, 8*(3), 91–107, https://doi.org/10.1300/J134v08n03_05

Dore, M. (2017). *Factors in the admissions process influencing persistence in a Master's of Science program in marine science* [Doctoral dissertation, Nova Southeastern University]. NSUWorks. https://nsuworks.nova.edu/cgi/viewcontent.cgi?article=1117&context=fse_etd

Easley, J., Strawderman, L., Babski-Reeves, K., Bullington, S., & Smith, B. (2021). Perceived quality factors in higher education. *Quality in Higher Education*, *27*(3), 306–323. https://doi.org/10.1080/13538322.2021.1909210

Educational Testing Service. (n.d.-a). *GRE® Guidelines for the Use of Scores*. https://www.ets.org/gre/institutions/admissions/using_scores/guidelines?WT.ac=40361_owt19_180820

Educational Testing Service. (n.d.-b). *Prepare for the GRE® General Test: Verbal Reasoning, Quantitative Reasoning, Analytical Writing*. https://www.ets.org/gre/revised_general/prepare/

Educational Testing Service. (n.d.-c). *Test Content*. https://www.ets.org/gre/revised_general/about/content

Educational Testing Service. (2019). *A Snapshot of the Individuals Who Took the GRE General Test July 2014–June 2019*. i–73. Educational Testing Service. https://www.ets.org/s/gre/pdf/snapshot_test_taker_data_2019.pdf

Educational Testing Service. (2021). *A Snapshot of the Individuals Who Took the GRE General Test July 2016–June 2021*. i–73. Educational Testing Service. https://www.ets.org/s/gre/pdf/snapshot.pdf

European Commission. (n.d.). *What is Erasmus+?* https://erasmus-plus.ec.europa.eu/about-erasmus/what-is-erasmus

Frasier, H. S. (2013). *An analysis of institutional characteristics that contribute to extended time to doctoral degree*. [Doctoral dissertation, University of Maryland]. *ProQuest LLC*. https://drum.lib.umd.edu/bitstream/handle/1903/14911/Frasier_umd_0117E_14734.pdf?sequence=1&isAllowed=y

Fu, Y. (2012). *The effectiveness of traditional admissions criteria in predicting college and graduate success for American and international students*. [Doctoral dissertation, University of Arizona]. ProQuest LLC. https://www.proquest.com/docview/917742486

Garces, L. M. (2014). Aligning diversity, quality, and equity: The implications of legal and public policy developments for promoting racial diversity in graduate studies. *American Journal of Education*, *120*(4), 457–480. https://doi.org/10.1086/676909

Gooding, Y. (2001). *The relationship between parental educational level and academic success of college freshmen* [Doctoral dissertation, Iowa State University]. Iowa State University digital repository. https://doi.org/10.31274/rtd-180813-12012

Gordon A. (2019). Do't Remind Me: Stereotype Threat in High-Stakes testing. *University of Baltimore Law Review, 48*(3), 387–412. https://scholarworks.law.ubalt.edu/cgi/viewcontent.cgi?article=2053&context=ublr

Hall, J. D., O'Connell, A. B., & Cook, J. G. (2017). Predictors of student productivity in biomedical graduate school applications. *PLOS ONE*, *12*(1), e0169121. https://doi.org/10.1371/journal.pone.0169121

Howell, L. L., Sorenson, C. D., & Jones, M. R. (2014). Are undergraduate GPA and general GRE percentiles valid predictors of student performance in an engineering graduate program? *International Journal of Engineering Education*, *30*(5), 1145–1165. https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=2342&context=facpub

Jamieson J. P. & Harkins S. G. (2009). The effect of stereotype threat on the solving of quantitative GRE problems: A mere effort interpretation. *Personality and Social Psychology Bulletin, 35*(10), 1301–1314. https://doi.org/ 10.1177/0146167209335165

King, M. R., Jennings, K., Chalkley, R. G., & Sealy, L. J. (2020). Questioning the value of the Graduate Record Examinations (GRE) in PhD admissions in biomedical engineering. *Annals of Biomedical Engineering*, *48*(8), 2155–2157. https://doi.org/10.1007/s10439-020-02552-7

Klieger, D. M., Belur, V., & Kotloff, L. J. (2017). Perceptions and uses of *GRE* ® scores after the launch of the *GRE* ® revised General test in August 2011. (Report No. ETS GRE®–17-01 ETS RR–17-03). Educational Testing Service. https://onlinelibrary.wiley.com/doi/full/10.1002/ets2.12130

Klieger, D. M., Cline, F. A., Holtzman, S. L., Minsky, J. L., & Lorenz, F. (2014). *New perspectives on the validity of the GRE ® General test for predicting graduate school grades: New perspectives for predicting graduate school grades*. (Report No. RR– 14-26). Educational Testing Service. https://doi.org/10.1002/ets2.12026

Kuncel, N. R., & Hezlett. S. A. (2007a). The utility of standardized tests: Response. *Science*, *316*, 1696–1697. https://doi.org/10.1126/science.316.5832.1694b

Kuncel, N. R., & Hezlett, S. A. (2007b). Standardized tests predict graduate students' success. *Science*, *315*(5815), 1080–1081. https://doi.org/10.1126/science.1136618

Kuncel, N. R., & Hezlett, S. A. (2010). Fact and fiction in cognitive ability testing for admissions and hiring decisions. *Current Directions in Psychological Science*, *19*(6), 339–345. https://doi.org/10.1177/0963721410389459

Kuncel, N. R., Wee, S., Serafin, L., & Hezlett, S. A. (2010). The validity of the Graduate Record Examination for master's and doctoral programs: A meta-analytic investigation. *Educational and Psychological Measurement*, *70*(2), 340–352. https://doi.org/10.1177/0013164409344508

Ledford, S., Monteagudo, M. M., Flores, A. N., Glass, J. B., & Cobb, K. M. (2020). Geogrexit: Strategies for removing the GRE as one barrier of entry to geoscience graduate programs. *Geological Society of America Abstracts with Programs, 52(6)*. https://doi.org/10.1130/abs/2020AM-353863

Lorden, J. F., Ed, Kuh, C. V., Ed, & Voytuk, J. A., Ed. (2011). *Research-doctorate programs in the biomedical sciences: Selected findings from the NRC assessment*. Washington, DC: The National Academies Press. https://doi.org/10.17226/13213.

Lott, J. L. I., Gardner, S., & Powers, D. A. (2009). Doctoral student attrition in the STEM fields: An exploratory event history analysis. *Journal of College Student Retention: Research, Theory and Practice*, *11*(2), 247–266. https://doi.org/10.2190/CS.11.2.e

Miller, E. M. (2019). Promoting student success in statistics courses by tapping diverse cognitive abilities. *Teaching of Psychology*, *46*(2), 140–145. https://doi.org/10.1177/0098628319834198

Miller, C., & Stassun, K. (2014). A test that fails. *Nature*, *510*, 303–304. https://www.nature.com/articles/nj7504-303a

Ministry of Education, Culture and Science of the Netherlands (2021). Regeling van de Minister van Onderwijs, Cultuur en Wetenschap 18 juni 2021, nr. 28415868, houdende andere bijdragen van studenten in het hoger onderwijs (Regeling andere bijdragen van studenten in het hoger onderwijs). https://zoek.officielebekendmakingen.nl/stcrt-2021-33378.html

Moneta-Koehler, L., Brown, A. M., Petrie, K. A., Evans, B. J., & Chalkley, R. (2017). The Limitations of the GRE in predicting success in biomedical graduate school. *PLOS ONE*, *12*(1), Article e0166742. https://doi.org/10.1371/journal.pone.0166742

Organization for Economic Co-operation and Development. (2017). *Scaling Procedures and Construct Validation of Context Questionnaire Data*. *PISA 2015 Technical Report*. https://www.oecd.org/pisa/sitedocument/PISA-2015-Technical-Report-Chapter-16-Procedures-and-Construct-Validation-of-Context-Questionnaire-Data.pdf

Payne, D. (2015). A common yardstick for graduate education in Europe. *Journal of the European Higher Education Area*, *2*, 21–48. https://www.handbook-internationalisation.com/en/handbuch/gliederung/?articleID=566#/Beitragsdetailansicht/675/566

Perez, K. (2011). *GRE as a Predictor of Graduate Student Success at a Hispanic Serving Institution of Higher Education* [Doctoral dissertation, Florida International University]. https://doi.org/10.25148/etd.FI11042602

Petersen, S. L., Erenrich, E. S., Levine, D. L., Vigoreaux, J., & Gile, K. (2018). Multi-institutional study of GRE scores as predictors of STEM PhD degree completion: GRE gets a low mark. *PloS One*, *13*(10), Article e0206570. https://doi.org/10.1371/journal.pone.0206570

Posselt, J. R. (2016). *Inside graduate admissions: Merit, diversity, and faculty gatekeeping*. Harvard University Press.

Postmes, L., Bouwmeester, R. A. M., de Kleijn, R. A. M., & van der Schaaf, M. (2022). Supervisors' untrained postgraduate rubric use for formative and summative purposes. *Assessment & Evaluation in Higher Education*, https://doi.org/10.1080/02602938.2021.2021390

Powers, D. E. (2017). Understanding the impact of special preparation for admissions tests. In R. E. Bennett & M. von Davier (Eds.), *Advancing Human Assessment*. *The Methodological, Psychological and Policy Contributions of ETS*. (pp. 553–564). Springer Open. https://doi.org/10.1007/978-3-319-58689-2

Rodríguez-Hernandez, C. F., Cascallar, E., & Kyndt, E. (2020). Socio-economic status and academic performance in higher education: A systematic review. *Educational Research Review*, *29*, Article 100305. https://doi.org/10.1016/j.edurev.2019.100305

Sackett P. R., Kuncel N.R., Arneson J. J., Cooper S. R., & Waters S. D. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance? *Psycholigical Bullitin, 135*(1), 1–22. https://doi.org/10.1037/a0013978

Sanford, R. (2009). *Predictors of international graduate student success in U.S. universities: analysis of noncognitive student variables and institutional characteristics*. [Doctoral dissertation, University of La Verne]. ProQuest LLC https://www.proquest.com/openview/0ca5804f25bd0ac8eeb599e839f8d656/1?pq-origsite=gscholar&cbl=18750&diss=y

Schmader T. & Johns M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*(3), 440–452. https://doi.org/ 10.1037/0022-3514.85.3.440

Schwager, I. T. L., Hülsheger, U. R., Bridgeman, B., & Lang, J. W. B. (2015). Graduate Student Selection: Graduate record examination, socioeconomic status, and undergraduate grade point average as predictors of study success in a western European University. *International Journal of Selection and Assessment, 23*(1), 71–79. https://doi.org/10.1111/ijsa.12096

Sealy, L., Saunders, C., Blume, J., & Chalkley, R. (2019). The GRE over the entire range of scores lacks predictive ability for PhD outcomes in the biomedical sciences. *PloS One*, *14*(3), Article e0201634. https://doi.org/10.1371/journal.pone.0201634

Skinner, M. (2018). *Student Mobility in the European Higher Education Area (EHEA)*. World Education News + Reviews. https://wenr.wes.org/2018/12/student-mobility-in-the-european-higher-education-area-ehea

Steele C. M. & Aronson J. (1995). Stereotype Threat and the Intellectual Test Performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797–811. https://doi.apa.org/doi/10.1037/0022-3514.69.5.797

Stricker L. J. & Rock D.A. (1993). Examinee background characteristics and GRE General test performance. (GRE Board Research Report No. 89-07R, ETS Research Report 92-80). Educational Testing Service. https://www.ets.org/research/policy_research_reports/publications/report/1993/ihle

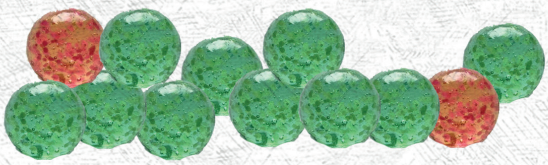
UC Office of the President. (2020). *University of California Board of Regents unanimously approved changes to standardized testing requirement for undergraduates*. https://www.universityofcalifornia.edu/press-room/university-california-board-regents-approves-changes-standardized-testing-requirement

Walters A. M., Lee S., & Trapani C. (2004). Stereotype threat, the test-center environment, and performance on the GRE General test. (Report No. 01-03R. ETS RR-04-37). Educational Testing Service. https://www.ets.org/research/policy_research_reports/publications/report/2004/hyzm

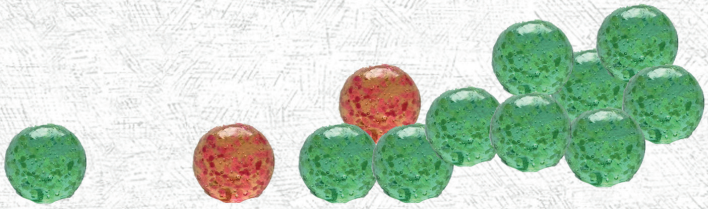
Zimmermann, J., von Davier, A. A., Buhmann, J. M., & Heinimann, H. R. (2017). Validity of GRE General test scores and TOEFL scores for graduate admissions to a technical university in Western Europe. *European Journal of Engineering Education*, *43*(1), 144–165. https://doi.org/10.1080/03043797.2017.1343277

Zwick, R. (2004). Is the SAT a "wealth test"? The link between educational achievement and socioeconomic status. In R. Zwick (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (pp. 203–216). RoutledgeFalmer. https://www.taylorfrancis.com/chapters/edit/10.4324/9780203463932-24/sat-wealth-test-link-educational-achievement-socioeconomic-status-rebecca-zwick

6



General Discussion



Most of the research on student selection focuses on admissions to undergraduate programs, while research into graduate selection is scarce. This dissertation builds on previous studies conducted on graduate selective admissions (but also considers the profound depth of knowledge on the undergraduate level). The main interest of this study was to enrich our understanding regarding student selection methods: primarily their predictive validity, but also other evaluative quality principles such as acceptability by stakeholders, cost-effectiveness, and possible procedural issues that should be accounted for in the implementation of selection methods (bias in testing, faking, etc.). Selective admissions to research masters' programs in STEM disciplines (and predominantly, life and natural sciences) within the Dutch higher education system (i.e., in research universities) were the focus of the studies in this thesis.

In this chapter, the main findings of this thesis will be discussed. Subsequently, we will outline the theoretical and practical contributions of the research as well as its limitations. Finally, we will address several relevant and complex themes that were outside the scope of this thesis but require attention for quality enhancement of selective admissions.

6.1 Summary of the Main Findings

The first aim of this thesis was to explore the available evidence on a variety of graduate selection methods in STEM fields. In our mapping review (Chapter 2), we found that selection methods such as prior grades, Graduate Record Examinations (GRE) General, intelligence assessments, and conscientiousness tests show a medium to strong extent of predictive validity toward certain dimensions of graduate study success. We also detected that selection methods such as letters of recommendation, tests on language proficiency, emotional intelligence, and need for cognition[23] show weak to medium extent of predictive validity. Importantly, among the examined selection methods, we also detected that some other selection methods—personal statements, traditional interviews, and personal traits such as extroversion and neuroticism—were shown to be invalid predictors. Prior research also emphasized that in addition to predictive validity, other considerations should be accounted for: procedural issues, acceptability by stakeholders, and cost-effectiveness. Therefore, we gathered available evidence on these considerations in relation to each selection method.

After conducting our review of prior research (Chapter 2), we then turned to our second aim of examining the practice of selective admissions (Chapter 3). Our study revealed what the current selective admissions practices to research masters' programs in life and natural sciences involve at a large Dutch research university. In this empirical examination, we distinguished between selection criteria (what admissions committees are looking for in applicants) and selection methods (how these criteria are measured).

23 Defined as an inclination to value activities that include effortful cognition.

We found that admissions committees assigned approximately equal importance (medium of size) to selection criteria related to grades, academic background of applicants, motivation, and research background. However, they gave less consideration to selection methods related to personality and personal competencies. Regarding the selection methods used for assessment of those selection criteria, we found that motivation letters, Curricula Vitae (CVs), recommendation letters, bachelor's transcripts, course contents, undergraduate Graduate Point Averages (UGPAs), grades for relevant courses, and type of prior educational institution were all extensively used as selection methods. We discussed the different extent of their susceptibility toward admissions biases and other procedural issues across the chapters.

Having detected common selection criteria and methods to Dutch research masters' programs in life and natural sciences (Chapter 3), we noted that there are selection methods that are not applied in admissions practices, even though they show potential value according to the existing scientific literature synthesized in our review (Chapter 2). We, therefore, set a goal to test the actual predictive validity of these selection methods in two empirical studies implemented in these STEM research masters' programs. We chose to focus on the selection methods: grade for bachelor's thesis (Chapter 4) and the GRE (Chapter 5). We found that both grade for bachelor's thesis and all three sections of the GRE have incremental validity in predicting graduate grade point average (GGPA). We also found that these two selection methods are predictive for other dimensions of graduate study success: Higher grade on bachelor's thesis is related to shorter time to degree, GRE-V is positively related to the supervisors' assessments on the internship rubric scale "performing research", and GRE-Q is positively related to supervisors' assessments on internship grade and internship rubrics on "practical skills" (marginally) and "structure and style" (marginally). Recognizing the pros and cons that implementation of these methods might involve, we discussed our recommendations for practitioners in the respective chapters and will give an overarching summary in this chapter.

6.2 Theoretical Contributions/Academic Relevance

This thesis extends our knowledge on evidence-based selection methods. It is the first study to systematically synthesize the existing research evidence on both cognitive and noncognitive selection methods at the graduate level on STEM disciplines. This is important to highlight because most prior reviews were conducted on the undergraduate level and often in the field of medical education (Cleland et al., 2012; Patterson et al., 2016; Stegers-Jager, 2018) with a different conceptualization of study success which included such fields as clinical performance. Our findings, therefore, integrate evidence on study success dimensions specific for STEM graduate education (e.g., research productivity).

Within the empirical part of this thesis, predictive value of different selection methods was examined in the context of Dutch STEM research masters' programs. What emerged from this study was that UGPA, bachelor thesis grade, and the scores on the three sections of the GRE showed meaningful prediction for study success. The limited predictive power of type of prior higher education institution (HEI) and prior field of study, despite a certain weight they carry in current admissions practice, was also demonstrated.

6.3 Practical Contributions/ Implications for Applicants & Practitioners

This thesis showed the complex nature of graduate student selection. Our findings indicate that there is no one objective or ideal selection method. Instead, uncertainty is unavoidable and should be managed appropriately. In practice, this would mean:

(a) prioritizing selection methods with predictive validity shown in scientific literature as well as avoiding reliance on a single method and using a variety of selection methods instead (e.g., UGPA and the GRE taken together explain more variance than only UGPA alone, standardized recommendation letters appear to be predictive of degree completion—a dimension of study success, which is difficult to predict using other selection methods),

(b) avoiding the usage of selection methods that have been proven systematically invalid in prior research as well in this study and/or can often enhance the risk of bias (e.g., traditional interview, personal statements, type of prior education institution, etc.),

(c) being aware of other quality evaluative principles, in addition to predictive validity (e.g., various procedural issues of selection methods, their acceptability, and cost-effectiveness),

(d) facilitating logistics and affordability of implementation of valid selection methods,

(e) creating the workflow in which the procedural issues such as admission bias, stereotype threat, differences in grading standards, grade inflation, coaching effects, and so forth are accounted for (e.g., admissions committees could be obligatorily trained to avoid admissions biases; standardized grade conversion system could be elaborated),

(f) ensuring transparency to applicants (so that applicants would accept a reject decision and do not start an appeal because they disagree with the selection procedure or feel they were treated unfairly),

(g) providing and adhering to guidelines on implementation of every selection method (e.g., not setting cut-off scores for standardized tests, see Chapter 5 for details).

The variety of selection methods which practitioners should consider including in their selective admissions to research masters' programs in STEM are as follows:

- undergraduate grade point average (UGPA)
- bachelor thesis grade
- GRE General (all three GRE scales—for prediction of GGPA, single GRE scales— for prediction of various internship-related dimensions of study success)
- language tests like TOEFL

With additional caution, the following methods could be considered:

- prior research experience (for admissions to research graduate programs)
- GPA for the last year of a bachelor's program
- standardized recommendation letters
- multiply mini-interviews
- standardized certified intelligence assessments

Inclusion of each of these selection methods should be guided by understanding which dimensions of study success these selection methods are capable of predicting, whether a selection method is accepted (and to what extent) by admissions committees and applicants, and whether the admissions committees are aware of the correct usage of a selection method.

From our review, it appeared that the selection methods that have no predictive value in graduate student selection are:

- personal statements
- traditional interviews
- narrative recommendation letters

Therefore, it is advised to avoid these instruments when making admissions decisions. This, however, does not mean that these instruments cannot be used for other purposes. For example, personal statements may be used for encouraging students to reflect on their motivation for a specific program and getting acquainted with it through exploration of the program's curriculum, internship opportunities, and career perspectives (Wouters et al., 2014).

6.4 Limitations

One of the limitations of this thesis is that for the empirical part, we did not have quantitative assessments of different selection methods allowing us to integrate these data in our statistical analyses (they were either in the format of written narratives such as recommendation letters and personal statements or admissions committees' notes after face-to-face or online interviews). This might have limited the statistical analysis for our research models to a certain extent. However, we did review the available research literature on those methods in graduate education.

The second limitation is that most of the data we used are the data of admitted students, which might question whether our results are generalizable toward the entire population of applicants to STEM research masters' programs. Even though our data would have been richer if we had had data on rejected applicants or applicants who had been admitted but never started their master's program, we do not think that our conclusions would be substantially different. In two out of three empirical studies, we were interested in establishing the relationships between the predictors (selection methods) and various dimensions of graduate study success. Given that our students were preselected, we have a restricted range on some of our predictors, which makes it harder to detect nonlinearities in their statistical effects. This, however, is unlikely to have affected the direction of the relationship. The data on rejected students are essential in studies aiming at establishing cut-off scores for admissions. This might be done, for example, using Signal Detection Theory; its application in admissions is a relatively new direction (van der Linden, 2018). All in all, we assume that the selection methods that showed predictive validity on a sample of selected students would be valid predictors for the entire population of applicants. The question arises then: How would the admitted population of students change if we use only selection methods, identified as valid in this dissertation and avoid using the ones which systematically did not show predictive validity? We hypothesize based on our findings that the population of admitted students would become more successful at least in terms of their performance on internships and their GGPA as well as more diverse in terms of their prior education (such as the type of prior HEI and field of studies) and perhaps also in terms of their socioeconomic status and related dimensions of diversity (e.g., migration background).

We do recognize that the limitation under consideration (the usage of data of mostly admitted students) could have limited the generalizability of our other empirical study in Chapter 3 as the number of rejected applicants who responded to the survey was small. In that study, we examined the perceptions of applicants on importance of selection criteria and methods in the admissions decisions that were made regarding them. It is possible that accepted applicants reported the overall process as transparent while the rejected applicants could have found it more challenging to navigate and we could not detect these challenges because of the low response rate from rejected applicants. We, therefore, chose to survey them after they knew their admissions results because they could report on the whole admissions process. Future studies could perhaps survey the applicants before they are informed about their admissions results. This approach would yield insights into both rejected and admitted applicants, however, researchers would need to account for the limitations arising from the earlier timing of sending the survey.

Another limitation of this dissertation is that the empirical part was conducted specifically in the Dutch higher education context, nevertheless, we indicate that certain generalizations are possible toward other European masters' programs. Since the Dutch

higher education system is a part of the European Higher Education Area (EHEA), the policy makers in other European countries possess the knowledge on similarities and differences between higher education systems and can determine to what extent our findings are generalizable to the other higher education contexts. We admit that the structure of at least the life sciences programs that were empirically examined in this thesis, is quite specific. However, there are research-intense programs (even though with various structure of components within the research-oriented curriculum) in other European countries as well. Also, several graduate study success dimensions, which we examined, are nearly universal for STEM masters' programs (i.e., GGPA, time to graduate degree, degree attainment, internship grade). Therefore, the generalizability to other higher education contexts is feasible.

6.5 Relevant Themes in the Field of Selective Admissions

There are several themes that were outside the scope of the studies in this thesis but are the topic of selective admissions and potentially exert strong effects on the population of graduate students. Each theme is briefly described below along with the relevant references for further reading.

6.5.1 Theme #1. Merit and Diversity

Focusing too much on merit, that is, on selection methods that cater to academic excellence and as such favor institutional gains, may compromise the diversity of the selected population (Wouters et al., 2018). In other words, student selection may lead to a more homogeneous student population. Therefore, one of the major themes that is deeply associated with the discourse on selective admissions is the question of how to form a diverse graduate student population, representative of societal composition in graduate education (in terms of gender, socioeconomic indicators, ethnicity/migrant background, etc.), while maintaining the aims of both merit and diversity (or, in other words, excellence and inclusiveness). Designing and implementing selection processes, where the adverse effects of admissions biases are minimized, is an important condition for creating a diverse population of students.

Another layer to this problem is the actual value attributed (or not) to the principle of student diversity. The academic judgements for selective programs are formed to a high extent by a notion of merit rather than diversity, even if the aim of diversity is declared. In her influential book, "How Professors Think: Inside the Curious World of Academic Judgment" (2009), Michèle Lamont showed that academic judgments on fellowships and research grants are guided by the notion of merit, while "diversity in particular can act as an additive, rather than as an alternative, standard of evaluation" (Lamont, 2009, p. 202). Lamont's study also showed that even in its additive value, academics prioritized

institutional and disciplinary diversity rather than ethno-racial, gender, or geographic diversity. These findings were confirmed in the context of graduate admissions. Julia Posselt (2014) in her work on admissions to elite selective graduate programs shows that the admissions committees of these programs usually have two rounds of selective admissions, and the notion of diversity only starts to play a role in the second round of reviews. To get to the second round, applicants have already been sorted in the first round by merit which is normally assessed by admissions committees according to conventional achievements on quantifiable metrics. Posselt notes that "Many of the students whose diversity contributions might have been considered assets had already been filtered from the pool" (2014, p. 507). She further explains that "the current two-tier review process relegates diversity to a secondary consideration, and it makes a standard of inclusive excellence conditional on conventional achievement" (2014, p. 507). These two works by Lamont and Posselt show that the meanings attributed to merit, diversity, and the decision-making process itself play a major role in how far organizations are from a culture of inclusive excellence: an integrated notion of merit and diversity, which is considered by some to be "critical to the wellbeing of democratic culture" (Association of American Colleges and Universities, 2014, as cited in Posselt, 2014, p. 48).

6.5.2 Theme #2. The Effects of Selection at Secondary School on Student Diversity in Higher Education

While the Dutch higher education system is perceived by many as egalitarian[24], the general principle of equity on entrance to higher education does not necessarily mean equality of opportunity (Reumer & van der Wende, 2010). In the Netherlands, educational opportunities become unequal when primary school ends after year 8 (around 12 years of age). At this point, pupils are selected for education into three differentiated (stratified) pathways: preuniversity education (6 yrs), senior general secondary education (5 yrs), and prevocational secondary education (4 yrs). The first two paths provide access to higher education, however, generally only preuniversity education provides direct access to university education. While there is a possibility for transition between paths, most students follow the tradition model. This means that as early as 12 years of age potentially successful students may miss out on their opportunity of a higher education degree and even more so for graduate level.

French sociologist Boudon (1974, as cited in Trueman, 2015) saw only two ways to eliminate inequality of educational opportunities: either to have an unstratified society or to make school systems undifferentiated. While the lack of stratification for any society is unlikely, the reorganization of a secondary school system seems more realistic to

24 This is due to (1) open access to higher education for students with the right type of secondary education qualification on most undergraduate programs (with a few exceptions such as medical and law schools, schools for fine and performing arts, and university colleges) and (2) approximately equal rankings of Dutch universities in international rankings.

implement. However, the idea of reorganizing secondary school education is often met with a lot of resistance from politicians, teachers, and parents (Becker & Hecken, 2009), also in the Netherlands, even after several attempts and repeated recommendations in this direction (Onderwijsraad, 2021). Presuming a prominent role of socioeconomic status and ethnicity in the secondary education stratification, we can assume that the pool of applicants to graduate education remains constrained in terms of its ethnic and socioeconomic diversity as many of those who could have participated have already been selected out.

6.5.3 Theme # 3. Self-Selection

Another theme, closely related to diversity, is self-selection: a strategy where applicants refrain from applying to programs/schools where their self-perceived chances of getting in are low. Some of the positive aspects of self-selection is that it reduces the chances of negative experiences during selective admissions such as being rejected. It also diminishes the workload for admissions committees and increases the acceptance rate (simply because there are fewer applicants). There are, however, negative aspects to self-selection. For example, students from low SES are more likely to self-select and are prone to strategic mistakes in school choice than students from high SES with the same academic achievements (Chen & Pereyra, 2019). The research indicates several reasons for this. For example, applicants from low SES (1) might find it more challenging to find and use effective application strategies to figure out which program fits them best, (2) might be less confident in their abilities and competencies than students from high SES, (3) have less family support as their parents are less involved in guiding them through educational pathways and less informed about (higher) educational pathways (Ajayi, 2013; Chen & Pereyra, 2019). For example, researchers, who examined effects of medical school selection on student motivation, indicate that applicants without parents in the medical profession may perceive that they are less equipped for selective admissions process, therefore, may abstain from applying to a medical school (Wouters et al., 2017). Overall, it appears that the effects of self-selection on the diversity of a student body are still relatively understudied in higher education (Meyer et al., 2019) and especially at the graduate level.

To diminish the effects of self-selection on diversity of student population, more research must be done on interventions that would prevent suitable candidates from not applying to their desired graduate programs. For example, graduate schools may want to gather more information on the study interests of their potential applicants before the applicants make their final study choices. Also, (graduate) schools could work on tweaking certain characteristics of their selective admissions process design that affect self-selection decisions. For example, from research on other educational selection fields (high schools, medical school), it appears that timing of application plays a major role in self-selection decisions (Chen & Pereyra, 2019). It is plausible to

assume that also the enhancement of various aspects within recruitment strategies could benefit the diversification of a student body (e.g., diversity of role models, using different communication styles, etc.).

6.5.4 Theme #4. Innovative Approaches to Student Selection

In the last five–ten years, new approaches for student selection have started emerging. Most of them are at an explorative stage or on a stage of pilots, have not gained widespread acceptance, and have only been explored at the undergraduate level. However, these approaches are innovative and promising. Therefore, some brief notes with corresponding references to full research are provided below.

Modular Approach. The modular approach breaks down a selective admissions process into components (Lievens & Corstjens, 2018). In this sense, it can be regarded as the opposite of a holistic approach in admissions. Lievens and Corstjens (2018) describe seven major components considered within the modular approach: (1) the stimulus format of a selection method: a written, pictorial, auditory, or audio-visual, live face-to-face interactions, etc.; (2) contextualization: whether a selection method contains realistic context which an applicant can refer to (and to what extent) or whether it is abstract; (3) stimulus presentation consistency or level of standardization across test takers; (4) modality of responses required from applicants (e.g., textual, pictorial, audio, videoconference, or face-to-face modality); (5) response evaluation consistency or level of standardization across admissions committees (it ranges from making free judgments to automated scoring); (6) primary individuals who respond to a selection method (e.g., an applicant themself, or their prior supervisor, or administration of their prior HEI); (7) how specific the instructions are (from weak instructions to specific instructions).

The general idea is that after breaking down a selection procedure by these components, the components can then be reconstructed in a wide assortment of selection procedures. Those in turn could be tested on their relevance, predictive validity, acceptability by stakeholders, etc. A major advantage of this approach is that researchers can examine the unique impact of each component on study success. In practice, this means that the whole selection procedure does not need to change drastically, but only certain components. For example, increasing standardization across evaluators or giving more specific instructions to applicants might already substantially increase valid predictions. More information on this approach is available in personnel selection (Lievens & Sackett, 2017) and healthcare education field (see for the review of studies Lievens & Corstjens, 2018).

Curriculum Sampling or Trial-Studying. Curriculum sampling is a procedure where applicants are presented with an assignment that is comparable to an assignment of their future education program (de Visser et al., 2017). Most of the existing research on this method has been undertaken at undergraduate level. An example of a typical assignment would be that a student is given course material to study (relevant to

their field) and then sits an exam on that material. The exam scores from these studies show high predictive validity and were positively perceived by applicants (Niessen et al., 2018; Niessen, 2018). To the best of our knowledge, this method has not been studied at the graduate level. Given the differences in complexity and curriculum of undergraduate and graduate programs, it would be interesting to see whether the findings on undergraduate level is also reflected at graduate level, and more specifically, how feasible and predictive it is toward graduate study success for research masters' programs.

Signal Detection Theory. Signal detection theory is a framework for conceptualizing performance on tasks that involve detecting a stimulus against a noisy background (Wixted, 2020). While it can be applied to several different fields, recently it has been used in the field of admissions to higher education (see van der Linden, 2018). From a SDT perspective, study success is a signal to be detected. In addition to informing admissions committees on *which* selection methods to use (what also regression analysis did in this dissertation), it also can inform on *how* to use them (because it allows to establish specific cut-off scores on selection methods that can be quantified). Even though it provides this additional value, one important condition applies. The establishment of cut-off scores is possible only if scores on dimensions of graduate study success are known for a complete sample of applicants. In other words, in order to establish cut-offs, the selection decisions must be made first hypothetically, and a full pool of applicants must enroll and study at an examined program. This condition is rarely possible to implement in practice as selective admissions imply that some candidates will be rejected. It is even more difficult to implement this condition on a regular basis, which would be desirable for conducting reevaluations/adjustments to cut-offs. Some research findings on SDT are available on an undergraduate level. There, an exceptional situation was possible, and the complete samples of applicants were available (van der Linden, 2018). Those findings have recently been extended to a graduate level and other higher education fields (Kurysheva et al., 2022).

Constrained Optimization. Constrained optimization (CO) is a mathematical procedure which allows to build models that ensure HEIs reach diversity goals while upholding excellence (Zwick, 2020). CO originated from operations research: an analytical technique of decision-making. A user can instruct the CO model to maximize an index that combines scores on evidence-based selection methods while imposing certain constraints. Users can choose one constraint or several. For example, it is possible to constrain a model based on the percentages of low SES students, minorities, undergraduate HEI, refugee status, geographical residency, etc.

This method has been used in several studies and there is one example of its practical implementation in Chile, where the CO model assisted in making high-quality and transparent admissions decisions (Duran & Wolf-Yadlin, 2011). The research shows that CO facilitates the admissions of a diverse pool of students while maintaining the

indexes of study success. What researchers do in their studies is to take a pool of admitted students and treat it as a pool of applicants. This ensures that the researchers have both admissions and graduate study success data. Then they hypothetically select a certain number of students using (1) the CO model in one hypothetical selection and (2) a student ranking procedure in other hypothetical selection (or different types of student rankings). Researchers then compare the background and academic characteristics of students "admitted" based on the CO model with those of students "admitted" based on rankings. For the CO model, researchers usually use a composite score of prior GPA and standardized tests scores while upholding the diversity constraints. For rankings, they use either the same composite score, or GPA only, or scores of standardized tests only. What they find is that constraints have little impact on study success of an admitted university year group. In other words, the diverse university year group, "admitted" through CO performs *at least* as good as university year groups, "admitted" by rankings. The university year group "admitted" by CO, however, has an advantage of meeting the specific diversity goals of a program as they were constrained from the beginning in the model. For more details on results of these studies, see Zwick and her team (2019, 2020). Overall, these findings are in line with the meta-analytical findings which show that mechanical data combination outperforms expert (clinical or holistic) data combination when predicting study success and multiple work criteria (Kuncel et al., 2013).

The complexity of the CO method is that it requires additional time and expertise to formulate the constraints in numeric terms as well as to integrate them into the CO model. There is obviously a myriad of diversity characteristics of a university year group that might be important to consider. The more desired characteristics, the more difficult it is to program the mathematical model. Apart from these technical challenges, imposing constrains based on ethnicity or socioeconomic status most likely will call into question the legality of such constraints in European countries since there is no legislation similar to affirmative action in the US. The decision would be rather bound to equal treatment legislation (e.g., Algemene wet gelijke behandeling, 2020; Charter of fundamental rights of the European Union, 2012).

6.5.5 Theme # 5. Alternative Admissions: Combining Selective Admissions with Lottery in the Ex Aequo Group ("Lottery in the Middle") or with "Lottery of the Qualified"

Earlier (in Chapter 1), we discussed lottery and weighted lottery as alternative ways of admissions. We also discussed the reasons why these procedures were abolished in the Netherlands in 2017 and, instead, selective admissions were introduced in programs with restricted numbers of study places. The same year, a proposal named *ex aequo* (lat. equal in rank) or "decentralized selection with lottery in the middle" was published (Visser, 2017). It was proposed to unite selective admissions with lottery in one procedure. The ex aequo way of admissions originates from the two assumptions. The

first assumption is that most applicants are in fact eligible for studying at a program of their choice and most of them will likely complete the program successfully. The second assumption is that the selection methods applied in selective admissions are not strong enough to produce reliable ranks: There are no means to reasonably argue that one applicant should be accepted while their next closest follower should be rejected.

What valid selection methods could do, however, is to assist in determining which applicants are predicted to be unsuccessful in studies to a high extent of certainty (unqualified candidates) and which applicants are predicted to be successful in their studies to a high extent of certainty (excellent candidates). Once these two decisions are made, the next step would be to apply a lottery toward "the middle": a group of applicants who appear to be suitable for the program, but who are not evidently excellent and, therefore, there is a certain extent of uncertainty regarding their future study success. This group can be substantial in size and is difficult to rank as scores show little variation. In this second step, these applicants "in the middle" receive equal ranks, and the fate decides who gets into the program of their choice and who does not. As Visser (2017) notes: "Fate has the great advantage that it does not look at socio-economic background, skin color, and the education of parents" (Visser, 2017). These factors that he mentions are all infiltrated to a certain extent in selective admissions decisions, especially in situations when admissions committees run out of means to distinguish the applicants reliably while being at the same time under pressure to decide in someone's favor.

The findings presented in this thesis speak in favor of combining decentralized selection with lottery in the ex aequo group. Even though this thesis showed that there are at least a few evidence-based selection methods with different extent of predictive validity, none of them (or their combination) predicts well enough to form a complete reliable rank of all applicants (in our empirical studies, the total explained variance was around 40% in graduate grade point average and for several of other dimensions of graduate success, even lower than that). However, these instruments should be sufficient for determining the two groups of excellent and unsuitable candidates in the first step of the proposed procedure.

A similar proposition was made by Michael Sandel in his book "The Tyranny of Merit" (2020); he names it a "lottery of the qualified". His proposal has one substantial difference and one addition compared to "the lottery in the middle". The difference is that Sandel suggests as the first step only to determine those who are predicted to be unsuccessful to a high extent of certainty (unqualified candidates). After they are winnowed out, the lottery of the rest (all qualified candidates), should be conducted. Therefore, even excellent applicants are included in the lottery. This is in line with the main point of his reasoning: we should treat merit "as a threshold qualification, not an ideal to be maximized" (Sandel, 2020, p. 185).

The addition that Sandel proposes to the described procedure is that each qualified applicant could get a different number of tickets depending on the aims that a selective higher education institution (HEI) pursues. For example, if the aims are to increase diversity, admit more first-generation students, and to counteract the transmission of higher social status, the HEI could give two or three lottery tickets to those applicants whose parents do not have a higher education degree.

In March 2021, a bill on decentralized lottery was sent to the Dutch House of Representatives. This bill allows for a combination of lottery and selection. By means of selection, only a portion of the candidates may be admitted directly, with the remaining candidates being drawn by lot. The possibility of admitting a group through selection, rejecting a group through selection, and drawing lots for the ex aequo group is therefore excluded. When it comes to graduate level, it can be argued that in the Dutch law "Wet op het hoger onderwijs en wetenschappelijk onderzoek" (2022; article 7.30b) there is space for interpretation that allows masters' programs to do so. This space is there, because for the graduate phase, it is not stipulated that selection must take place exclusively on the basis of two qualitative selection criteria. Thus, after applying at least two qualitative admissions criteria to a full pool of applicants with completed applications, a lottery can be held for the ex aequo group (van den Hoeven, personal communication, 18 May, 2022). Future research on this alternative admissions method could provide more insights on (dis)advantages of this procedure.

6.6 Closure

We explored a field of selective admissions: a topic of high societal relevance. The implications of this study's findings have resulted in the following recommendations. First, we should commit ourselves to evaluating all applied selection methods against scientific evidence. Second, we should stop using selection methods that lack predictive validity and we should become aware of the procedural issues involved in implementation of selection methods with acceptable predictive validity. Third, we should integrate evidence-based selection methods in a sound admissions process where (1) self-selection effects are accounted for as much as possible, (2) admissions decisions are checked for admissions biases and regularly evaluated based on data, (3) admissions process is open for scientifically-sound adjustments. On a final note, we should also keep an open mind to alternative ways of admissions rather than pure selective admissions in the current form as they may present us with new opportunities to make fairer, more objective, and transparent admissions decisions.

6.7 References

Ajayi, K. F. (2014*). School choice and educational mobility: Lessons from secondary school applications in Ghana* [Working paper]. Department of Economics, University of California-Berkeley.

Algemene wet gelijke behandeling [The Dutch General Equal Treatment Act] (2020). https://wetten.overheid.nl/BWBR0006502/2020-01-01/0/afdrukken

Becker, R., & Hecken A. E. (2009). Why are working-class children diverted from universities?—An empirical assessment of the diversion thesis. *European Sociological Review, 25*(2), 233–250. https://www.jstor.org/stable/25548320

Charter of Fundamental Rights of the European Union (2012). 2012/C 326/02. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT

Chen, L., & Pereyra, S. J., (2019). Self-selection in school choice. *Games and Economic Behavior, 117*, 59–81. https://doi.org/10.1016/j.geb.2019.06.004

Cleland, P. J., Dowell, D. J., McLachlan, P. J., Nicholson, D. S., & Patterson, P. F. (2012). *Identifying best practice in the selection of medical students (literature review and interview survey)*. University of Aberdeen. University of Dundee. Queen Mary University of London. Working Psychology Group. https://www.gmc-uk.org/-/media/gmc-site-images/about/identifyingbestpracticeintheselectionofmedicalstudentspdf51119804.pdf?la=en&hash=D06B62AD514BE4C3454DEECA28A7B70FDA828715

de Visser, M., Fluit, C., Fransen, J., Latijnhouwers, M., Cohen-Schotanus, J., & Laan, R. F. J. (2017). The effect of curriculum sample selection for medical school. *Advances in Health Sciences Education, 22*(1), 43-56. https://doi.org/10.1007/s10459-016-9681-x

Duran, G. & Wolf-Yadlin, R. (2011). A mathematical programming approach to applicant selection for a degree program based on affirmative action. *Interfaces, 41*, 278–288. https://www.jstor.org/stable/23016305

Higher education and scientific research Act of the Netherlands [Wet op het hoger onderwijs en wetenschappelijk onderzoek], § 7.53 (2017). https://zoek.officielebekendmakingen.nl/kst-35765-2.html

Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: a meta-analysis. *Journal of applied psychology, 98*(6), 1060–1072. https://doi.apa.org/doi/10.1037/a0034156

Kurysheva, A., van Ooijen-van der Linden, L., van der Smagt, M., & Dilaver, G. (2022). potential and unique incremental value of signal detection theory as a research method in higher education: An example from selective graduate admissions [Manuscript submitted for publication]. Center of Education and Training, University Medical Center Utrecht & Faculty of Social and Behavioural Sciences, Utrecht University.

Lamont, M. (2009). *How professors think: Inside the curious world of academic judgment*. Harvard University Press.

Lievens, F. & Corstjens, J. (2018). New approaches to selection system design in healthcare: The practical and theoretical relevance of a modular approach. In F. Patterson & L. Zibarras (Eds.), *Selection and recruitment in the healthcare professions. Research, Theory and Practice (*pp. 167–192). Palgrave Macmillan.

Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology, 91*(5), 1181–1188. https://doi.org/10.1037/0021-9010.91.5.1181

Meyer, H., Zimmermann, S., Hissbach, J., Klusmann D., & Hampe W. (2019). Selection and academic success of medical students in Hamburg, Germany. *BMC Med Educ*. Article 19: 23. https://doi.org/10.1186/s12909-018-1443-4

Niessen, S. (2018). *New rules, new tools: Predicting academic achievement in college admissions* [Doctoral dissertation, University of Groningen]. Rijksuniversiteit Groningen. https://research.rug.nl/en/publications/new-rules-new-tools-predicting-academic-achievement-in-college-ad

Niessen A. S. M., Meijer R. R., Tendeiro J. N. (2018) Admission testing for higher education: A multi-cohort study on the validity of high-fidelity curriculum-sampling tests. *PLoS ONE 13*(6), Article e0198746. https://doi.org/10.1371/journal.pone.0198746

Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., & Cleland, J. (2016). How effective are selection methods in medical education? A systematic review. *Medical Education*, *50*(1), 36–60. https://doi.

org/10.1111/medu.12817

Posselt, J. R. (2016). *Inside graduate admissions: Merit, diversity, and faculty gatekeeping*. Harvard University Press.

Posselt, J. R. (2014). Toward inclusive excellence in graduate education: Constructing merit and diversity in PhD admissions. *American Journal of Education*, *120*(4), 481–514. https://doi.org/10.1086/676910

Reumer, C., & van der Wende, M. (2010). Excellence and diversity: The emergence of selective admissions policies in Dutch higher education. A case study on Amsterdam University College. Research & Occasional Paper Series: CSHE.15.10, 1–28. https://cshe.berkeley.edu/publications/excellence-and-diversity-emergence-selective-admission-policies-dutch-higher-0

Sandel, M. J. (2020). *The tyranny of merit. What's become of the common good?* Farrar, Straus and Giroux.

Stegers-Jager, K. M. (2018). Lessons learned from 15 years of non-grades-based selection for medical school. *Medical Education*, *52*(1), 86–95. https://doi.org/10.1111/medu.13462

Trueman, C. N. (2015). Raymond Boudon. Retrieved March 23, 2022, from https://www.historylearningsite.co.uk/sociology/education-and-sociology/raymond-boudon/

van der Linden, L. (2018). *Prediction of study success - creation of magic zones* [Doctoral dissertation, Utrecht University]. https://www.narcis.nl/publication/RecordID/oai:dspace.library.uu.nl:1874%2F373433

Visser K. (2017, August 23). *Tilt the selection in higher education [Kantel de selectie in het hoger onderwijs]* https://www.scienceguide.nl/2017/08/kantel-de-selectie-in-het-hoger-onderwijs/

Wixted, J. T. (2020). The Forgotten History of Signal Detection Theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(2), 201–233. https://doi.org/10.1037/xlm0000732

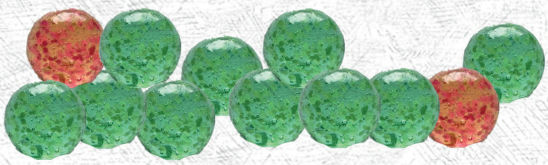
Wouters, A., Bakker, A. H., van Wijk, I. J., Croiset, G., & Kusurkar, R. A. (2014). A qualitative analysis of statements on motivation of applicants for medical school. *BMC Med Educ 14*, Article 200. https://doi.org/10.1186/1472-6920-14-200

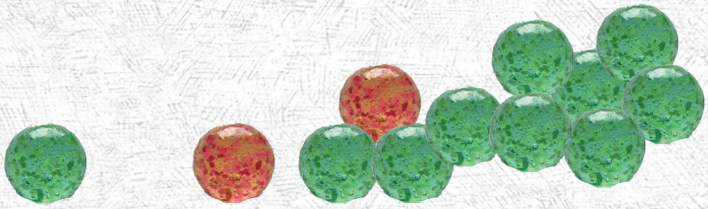
Wouters, A., Croiset, G., Isik, U., & Kusurkar R. A. (2017). Motivation of Dutch high school students from various backgrounds for applying to study medicine: a qualitative study. *BMJ Open* .Article 7:e014779. https://doi.org/10.1136/bmjopen-2016-014779

Wouters, A., Croiset, G., & Kusurkar, R.A. (2018). Selection and lottery in medical school admissions: Who gains and who loses? *MedEdPublish NaN*, Article 7:271. https://doi.org/10.15694/mep.2018.0000271.1

Zwick, R. (2020). Using mathematical models to improve access to postsecondary education. In M. E. Oliveri & C. Wendler (Eds.), *Higher Education Admissions Practices*. *An International Perspective* (pp. 333–346). Cambridge University Press.

Zwick, R., Ye, L., & Isham, S. (2019). Using constrained optimization to increase the representation of students from low-income neighborhoods. *Applied Measurement in Education 32*(4), 281-297. https://doi.org/10.1080/08957347.2019.1660346





English summary

The Dutch research universities are increasingly recognizing the need for evidence-based selection methods for admissions to research masters' programs. This need is becoming urgent due to several reasons: (1) a switch from an open admissions model and the weighted lottery approach to selective admissions, (2) yearly increasing numbers of (inter)national applicants, (3) rising diversification of application files and a commitment of universities to the diversity mission, and (4) societal expectations for fair, objective, inclusive, and transparent admissions. This doctoral dissertation addresses the multiple considerations regarding selection methods: predictive validity, acceptability by stakeholders, procedural issues, and transparency. The ultimate goals are: (a) to contribute to the research field of student selection and (b) to help admissions committees and other involved practitioners to apply evidence-based selection methods and justify the selection decisions they make.

Chapter 1. The Introduction provides the rationale for this doctoral dissertation. The chapter begins with a description of the Dutch Higher Education context and includes a brief overview of the history of admissions in the Netherlands, the effects of the Bologna Process, the increasing internationalization, and equity considerations on selective admissions in Dutch higher education.

The chapter then focuses on defining the scope of this doctoral dissertation. First, theoretical and methodological perspectives on student selection are outlined, as well as the urgency of conducting the proposed studies. Student selection cannot take place without answering the question: "What is graduate study success?" In this thesis, it is defined through various dimensions such as degree attainment, undergraduate average grade, time to degree, performance on research-related tasks, and some others. Additionally, the purpose, research questions, and the overarching structure of this dissertation are described.

Chapter 2. This chapter is the first study in our exploration of the field of graduate selective admissions. It presents a review of existing literature over the last 15 years and identifies the gaps where future research should focus. The research questions of this chapter are: What evidence is provided in the research literature on the extent to what different selection methods in graduate admissions in Science, Technology, Engineering and Mathematics (STEM) fields are valid, reliable, accepted by the stakeholders, and cost-effective? What are the procedural issues of the existing selection methods in graduate admissions?

To answer these questions, a systematic search of research literature was conducted. The graduate selection methods identified in the literature were classified into ten categories: (1) prior grades, (2) standardized testing of academic abilities, (3) letters

of recommendation, (4) interviews, (5) personal statements (i.e., motivation letters), (6) personality assessments, (7) intelligence assessments, (8) language proficiency, (9) prior research experience, and (10) various, rarely studied selection methods that do not fall under more common methods above (such as resumes, selectivity of prior higher education institution (HEI), former (type of) HEI, amount and quality of research experience, or composite scores). Each of these ten categories was assessed against four evaluative quality principles: predictive validity (and reliability, when available), acceptability by stakeholders, procedural issues, and cost-effectiveness.

The systematic literature search and the latter screening delivered 80 studies, which were included in the review and yielded the following results: the predictive validity of selection methods toward several study success dimensions varies from negligible extent (e.g., personal statements, traditional interviews, and personal traits such as extroversion and neuroticism) through weak-to-medium extent (letters of recommendation, tests on language proficiency, emotional intelligence, and need for cognition) to medium-to-strong extent (prior grades, Graduate Record Examinations (GRE) General, intelligence assessments, and conscientiousness). There were also selection methods that require more evidence to draw conclusions on their predictive value, even though a few published studies on these methods show their value in this regard (namely, prior research experience, multiple mini-interviews, and selectivity of prior institution). Each category of selection methods has their specific procedural issues, which we described based on the literature. The acceptance of various selection methods also varies greatly. For example, mini interviews are almost never applied in selection for STEM graduate programs, while undergraduate grade point average (UGPA) is applied widely. Likewise, the reactions of applicants toward different selection methods vary. The interested reader can refer to the Appendix at the end of this book to see the summaries of each study, included in this review.

Based on the findings of this review, the recommendations for practitioners were formulated. The advice given concerns excluding invalid selection instruments from admissions requirements, discussing, and deciding beforehand what dimensions of study success are most important for a given master's program (and choosing the selection methods according to these dimensions), using methods with a proven predictive validity, accounting for acceptability considerations and procedural issues of applied selection methods, ensuring the accountability of the admissions process.

Chapter 3. In this chapter, an examination of the currently applied selective admissions criteria and methods of two-year research masters' programs in the Graduate Schools of Life Sciences and Natural Sciences of Utrecht University (the Netherlands) is undertaken. Moreover, the transparency of the selection criteria toward applicants is evaluated.

For this study, two versions of a survey were designed: one for members of admissions committees and another for university applicants. Based on the existing literature and the admissions requirements found on university webpages, 51 admissions criteria were formulated, which were included in this survey. Two types of analyses were conducted in order to explore the importance of these criteria in admissions decisions. The first type of analysis was the ranking of all 51 individual admissions criteria from highest to lowest. The second type of analysis was performed by categorization the 51 criteria into domains. Six domains emerged and were labeled as (1) academic background, (2) grades, (3) cognitive ability, (4) research background, (5) personality and personal competencies, and (6) motivation factors. The results of two versions of the survey were compared, meaning that the perceptions of applicants on importance of admissions criteria were contrasted with the importance of admissions criteria, as reported by admissions staff members.

What was found is that the criteria related to applicants' academic background, grades, motivation factors, but also cognitive ability and research background, played an important role in admissions decisions. However, according to admissions committees, the criteria related to personality and personal competencies, did not play an important role in admissions decisions. When it comes to assessment of selection criteria by applying selection methods, an unstructured motivation letter, a CV, and a transcript with grades are among the most frequently used selection methods. This study also highlighted that the extent of transparency of selection criteria to applicants was moderate.

Overall, this study showed that admissions committees use selection criteria and methods with and without proven predictive value for graduate study success. Moreover, some of the used selection methods are prone to admissions biases. Admissions committees are advised to only apply selection criteria and methods that are evidence-based, resistant to admissions biases, and transparent for applicants.

Chapter 4. This chapter focuses on exploring the predictive value of information on applicants' prior education and performance, available in institutional data. In this study, these data are called undergraduate academic indicators and the advantage of this data is that it increases the possible generalizability of study findings, because the examined indicators are readily available at most higher education institutions (HEIs). Four undergraduate academic indicators were the focus of this study: UGPA, undergraduate thesis grade as an operationalization of research experience, type of prior HEI, and prior field of study. The relationship between undergraduate indicators and graduate study success (defined as graduate degree completion, GGPA, and graduate time to degree) was examined.

To examine this relationship, multilevel regression analyses was used on a sample of research masters' students in the field of life sciences (N = 1,792). What was found was that UGPA, as well as undergraduate thesis grade are valid predictors of graduate grade

point average (GGPA). Type of prior HEI was also shown to be predictive toward GGPA, but to a small extent, and we specifically discuss that this criterion should not be applied in selection of students to graduate programs. Regarding two other dimensions of graduate study success (degree completion and time to degree), the evidence showed that the undergraduate academic indicators that were used are predictive of them only to a small extent. It is possible that because the aspects of the students' program and their experiences on the program play a more prominent role in determining degree completion and time to degree then the undergraduate academic indicators available upon admissions. Even though our findings on the predictive value of thesis grade (in addition to UGPA) can be considered by admissions committees in order to implement them in practice, the admissions committees should account for the context in which these grades were obtained. In this regard, the models, alternative to the model of meritocratic equality of opportunity, can provide the admissions practitioners with relevant insights.

Chapter 5. This chapter examines whether the Graduate Record Examinations (GRE) has predictive value toward various dimensions of graduate study success on English-taught STEM programs at a large European university. The GRE is a standardized test, which consists of three scales: verbal reasoning, quantitative reasoning, and analytical writing. The advantage of standardized testing is that it helps with the comparison of application files because the GRE scores are on the same scale for all applicant (national and educational) backgrounds.

To determine whether the GRE is potentially useful for admissions to English-taught research masters' programs in Europe, data from 167 starting masters' students were collected. First, the starting masters' students took the GRE General and provided information on their previous study success, educational background, and demographical characteristics. To analyze these data, the hierarchical regression analyses were conducted. The results from this study showed that all GRE scales predicted Graduate Grade Point Average. Individual GRE scales predicted internship grade and supervisors' assessments of students' skills on performing research, their practical skills, and structure and style of their research reports. None of the GRE scales predicted supervisors' assessments of content of students' research reports. The identified relationships held after accounting for socioeconomic status. It was concluded that the GRE is a reasonable predictor of graduate study success for English-taught STEM research masters' programs. The advantages and disadvantages of the GRE application in selective admissions, as well as their legal limitations were addressed.

Chapter 6. The final chapter of this dissertation provides a summary of the main findings of this doctoral dissertation. In this chapter, the theoretical and practical contributions of this research as well as its limitations are described. Admissions practitioners may refer to the section titled "Practical Contributions/ Implications for

Applicants & Practitioners" of this chapter to gain insights, guidelines, and possibly inspiration for their daily work.

The rest of the chapter is dedicated to pinpointing several important themes that were not addressed in detail in this thesis (because they were outside of its scope). However, these themes constitute an important discourse within research and practice of selective admissions. The themes are as follows: (1) merit and diversity; (2) the effects of selection at secondary school on student diversity in higher education; (3) the pros and cons of self-selection; (4) innovative approaches to student selection such as modular approach, curriculum sampling or trial-studying, signal detection theory, and constrained optimization; (5) alternative model of selective admissions: combining selective admissions with lottery in the ex aequo group ("lottery in the middle") or with "lottery of the qualified". The discussion of these broad themes as well as the findings from the studies presented in this dissertation lead to the conclusion that the model of pure selective admissions in its current form might not be the best model to serve to the goals of fairness, objectiveness, and transparency in admissions. Admissions practitioners, therefore, should remain open to alternative ways of admissions and researchers should continue questioning each of those models for the ultimate goal of achieving equity and serving the growing diverse student community in the best possible way.

Nederlandse samenvatting

De Nederlandse universiteiten erkennen steeds meer de noodzaak van evidence-based (wetenschappelijk onderbouwde) selectiemethoden voor toelating tot research masterprogramma's. Evidence-based selectie wordt steeds urgenter om verschillende redenen: (1) de omschakeling van het open toelatingsmodel en de gewogen loterij naar selectie, (2) het jaarlijks stijgend aantal (inter)nationale kandidaten, (3) de groeiende diversiteit van aanmeldingen in combinatie met de toewijding van universiteiten aan hun diversiteitsmissie en (4) de maatschappelijke verwachting op eerlijke, objectieve, inclusieve en transparante toelating. Dit proefschrift bespreekt verschillende overwegingen rondom selectiemethoden: voorspellende validiteit, acceptatie door belanghebbenden, procedurele kwesties en transparantie. De doelen van dit proefschrift zijn: (a) bijdragen aan het onderzoeksveld van studentenselectie en (b) toelatingscommissies en andere betrokken beroepsbeoefenaren helpen om evidence-based selectiemethoden toe te passen.

Hoofdstuk 1. De inleiding zet de basale redenatie van dit proefschrift uiteen. Het hoofdstuk begint met een beschrijving van de Nederlandse context van het hoger onderwijs (een kort overzicht van de geschiedenis van toelatingen in Nederland; de gevolgen die het Bolognaproces en de toenemende internationalisering hadden op selectieve toelatingen; rechtvaardigheidsoverwegingen in het Nederlandse hoger onderwijs).

Het centrale deel van dit hoofdstuk is gewijd aan het definiëren van de reikwijdte van dit proefschrift. Eerst worden theoretische en methodologische perspectieven op de selectie van studenten geschetst, evenals de urgentie van het uitvoeren van de voorgestelde onderzoeken. Vervolgens wordt de aandacht gevestigd op het feit dat studentenselectie niet kan plaatsvinden zonder het antwoord op de vragen: "Waarvoor selecteren we eigenlijk studenten? Wat is studiesucces?" Studiesucces wordt daarom gedefinieerd middels verschillende dimensies die in de opvolgende hoofdstukken worden toegepast. Aan het einde van de inleiding worden de doelstellingen vermeld en de onderzoeksvragen samen met de overkoepelende structuur van dit proefschrift beschreven.

Hoofdstuk 2. Dit is de eerste steen in onze verkenning van het veld van selectieve toelating van kandidaten. Dit hoofdstuk geeft een overzicht van bestaande literatuur, brengt de kennis die in de afgelopen 15 jaar is opgebouwd in kaart en identificeert de hiaten die toekomstig onderzoek zou moeten opvullen. De onderzoeksvragen van dit hoofdstuk zijn: Welk bewijs bestaat in de onderzoeksliteratuur over validiteit, acceptatie, kosteneffectiviteit, en transparantie van verschillende selectiemethoden in Science, Technology, Engineering and Mathematics (STEM) - gebieden? Wat zijn de procedurele problemen van de bestaande selectiemethoden?

Om deze vragen te beantwoorden is er een systematische zoektocht naar onderzoeksliteratuur gedaan. De selectiemethoden, geïdentificeerd in de literatuur, werden onderverdeeld in tien categorieën: (1) eerdere cijfers, (2) gestandaardiseerde toetsing van academische vaardigheden, (3) aanbevelingsbrieven, (4) interviews, (5) persoonlijke verklaringen (bijv. motivatiebrieven), (6) persoonlijkheidsbeoordelingen, (7) intelligentiebeoordelingen, (8) taalvaardigheid, (9) eerdere onderzoekservaring, en (10) verschillende, zelden bestudeerde selectiemethoden die niet onder de meer gebruikelijke methoden hierboven vallen (zoals cv's, selectiviteit van een eerdere instelling voor hoger onderwijs (IHO), voormalige (type) IHO, hoeveelheid en kwaliteit van onderzoekservaring, of samengestelde scores). Elk van deze tien categorieën werd beoordeeld door middel van vier evaluatieve kwaliteitsprincipes: voorspellende validiteit (en betrouwbaarheid, indien beschikbaar), acceptatie door belanghebbenden, procedurele problemen en kosteneffectiviteit.

Het systematisch literatuuronderzoek en de screening leverde 80 studies op, die we in de review hebben opgenomen. De resultaten van dit literatuuronderzoek zijn als volgt: de voorspellende validiteit van selectiemethoden voor verschillende dimensies van studiesucces varieert van verwaarloosbare mate (bijv. persoonlijke uitspraken, traditionele interviews en persoonlijke kenmerken zoals extraversie en neuroticisme) tot zwak- middelmatig mate (aanbevelingsbrieven, tests op taalvaardigheid, emotionele intelligentie en behoefte aan cognitie) tot middelmatig-sterke mate (vroegere cijfers, Graduate Record Examinations (GRE) General, intelligentiebeoordelingen en consciëntieusheid/zorgvuldigheid). Er zijn ook selectiemethoden waarvoor meer bewijs nodig is om conclusies te trekken over hun voorspellende waarde, hoewel de eerste gepubliceerde studies over deze methoden hun bruikbaarheid in dit opzicht aantonen (namelijk eerdere onderzoekservaring, meerdere mini-interviews en selectiviteit van eerdere instellingen). Elke categorie selectiemethoden heeft zijn specifieke procedurele problemen, die we hebben beschreven op basis van de literatuur. De acceptatie van verschillende selectiemethoden varieert ook sterk: mini-interviews worden bijvoorbeeld bijna nooit toegepast bij selectie voor STEM-graduate-programma's, terwijl Undergraduate Grade Point Average (UGPA) veel wordt toegepast. Evenzo variëren de reacties van kandidaten op verschillende selectiemethoden, hoewel het meeste onderzoek op dit gebied zich afspeelde binnen personeelsselectie (en niet studentenselectie). De geïnteresseerde lezer kan de bijlage aan het einde van dit boek raadplegen om kerninformatie te krijgen over alle onderzoeken (en hun resultaten), die in dit overzicht zijn opgenomen.

Op basis van de bevindingen van deze review worden de aanbevelingen voor beoefenaars geformuleerd. Het advies is: uitsluiten van gebrekkelijke selectie-instrumenten van toelatingseisen; vooraf bespreken en beslissen welke dimensies van studiesucces belangrijkst zijn bij een bepaalde masteropleiding (en kiezen van de selectiemethoden op basis van deze dimensies); gebruik maken van methoden met

een bewezen voorspellende validiteit; rekening houden met acceptatie en procedurele kwesties van de selectiemethoden.

Hoofdstuk 3. In dit hoofdstuk is een onderzoek gedaan naar de huidige selectieve toelatingscriteria en methoden van tweejarige researchmasterprogramma's in de Graduate Schools of Life Sciences en Natural Sciences van de Universiteit Utrecht. Daarnasst wordt de transparantie van de selectiecriteria naar kandidaten beoordeeld.

We hebben een enquête ontworpen in twee versies: de ene versie is verzonden naar leden van toelatingscommissies en de andere versie is verzonden naar kandidaten. Op basis van de bestaande literatuur en de webpagina's met toelatingseisen hebben we 51 toelatingscriteria geformuleerd die in dit onderzoek zijn opgenomen. Er zijn twee soorten analyses uitgevoerd om het belang van deze criteria bij toelatingsbeslissingen te onderzoeken. Het eerste type analyse was een rangschikking van alle 51 individuele toelatingscriteria: we bespreken vervolgens welke criteria de hoogste posities innemen en welke de laagste posities innemen. Het tweede type analyse werd uitgevoerd door deze 51 criteria in domeinen te categoriseren. Zes domeinen kwamen naar voren en we labelden ze als (1) academische achtergrond, (2) cijfers, (3) cognitieve vaardigheden, (4) onderzoek achtergrond, (5) persoonlijkheid en persoonlijke competenties, en (6) motivatiefactoren. Omdat we ook transparantie wilden evalueren, hebben we de resultaten van twee versies van de enquête tegenover elkaar gezet. We hebben dus de perceptie van kandidaten vergeleken met het werkelijke belang van toelatingscriteria, zoals gerapporteerd door toelatingscommissies.

Wat we ontdekten is dat criteria met betrekking tot de academische achtergrond, cijfers, motivatiefactoren, maar ook cognitieve vaardigheden en onderzoek achtergrond van de kandidaten een belangrijke rol speelden bij toelatingsbeslissingen. Volgens toelatingscommissies speelden de criteria met betrekking tot persoonlijkheid en persoonlijke competenties echter geen belangrijke rol bij toelatingsbeslissingen. Een andere bevinding houdt verband met de selectiemethoden die worden gebruikt voor de beoordeling van selectiecriteria. We vonden dat een ongestructureerde motivatiebrief, een cv en een transcript met cijfers het vaakst worden gebruikt, hoewel ook andere selectiemethoden vaak worden toegepast. De mate van transparantie van de selectiecriteria voor kandidaten was matig.

Al met al blijkt uit dit onderzoek dat toelatingscommissies selectiecriteria en -methoden hanteren met en zonder bewezen voorspellende waarde voor het studiesucces. Bovendien zijn sommige van de gebruikte selectiemethoden vatbaar voor vooroordelen over toelating. Op basis van deze gevolgtrekking roepen we toelatingscommissies op om alleen die selectiecriteria en selectiemethoden gebruiken die op feiten zijn gebaseerd, bestand zijn tegen vooroordelen rondom de toelating ("admissions biases") en transparant zijn voor kandidaten.

Hoofdstuk 4. Hoofdstuk 4 richt zich op het verkennen van de voorspellende waarde van informatie over de vooropleiding en prestaties van kandidaten, beschikbaar in institutionele gegevens. In dit onderzoek worden deze gegevens "undergraduate academische indicatoren" genoemd. Het voordeel van deze gegevens is dat het de mogelijke generaliseerbaarheid van onderzoeksresultaten vergroot, omdat de onderzochte indicatoren direct beschikbaar zijn bij de meeste hogeronderwijsinstellingen (HOI's). We hebben ons gericht op vier undergraduate academische indicatoren: undergraduate grade point average (UGPA), scriptiecijfer als een operationalisering van onderzoekservaring, type eerdere HOI en eerder vakgebied. We onderzochten hun relatie met graduate studiesucces (gedefinieerd als graduate degree-voltooiing, graduate grade point average (GGPA) en graduate time to degree).

Om deze relatie te onderzoeken is gebruik gemaakt van multilevel regressieanalyses op een sample van studenten van research masters op het gebied van life sciences (N = 1.792). We ontdekten dat UGPA en het scriptiecijfer valide voorspellers zijn van het gemiddelde van GGPA. Het type eerdere HOI bleek ook voorspellend te zijn voor GGPA, maar in klein mate. We betogen dat dit criterium niet moet worden toegepast bij de selectie van studenten voor graduate programma's. Met betrekking tot twee andere dimensies van graduate studiesucces (voltooiing en doorlooptijd), hebben we laten zien dat de academische indicatoren die we hebben gebruikt slechts in kleine mate voorspellend zijn: mogelijk omdat de aspecten van het programma zelf en de ervaringen van studenten tijdens het programma een meer prominente rol spelen bij het bepalen van deze twee dimensies dan undergraduate academische indicatoren die beschikbaar zijn bij toelating. De bevindingen over de voorspellende waarde van het scriptiecijfer (naast UGPA), kunnen door toelatingscommissies worden meegenomen om ze in de praktijk te brengen. We bespreken echter ook dat de toelatingscommissies bij het nemen van selectiebeslissingen op basis van cijfers rekening moeten houden met de context waarin deze cijfers zijn behaald. In dit verband vestigen we de aandacht op de modellen, alternatief voor het model van meritocratische gelijkheid van kansen.

Hoofdstuk 5. Dit hoofdstuk richt zich op de vraag of de Graduate Record Examinations (GRE) voorspellende waarde hebben voor verschillende dimensies van graduate studiesucces op Engelstalige STEM-programma's aan een grote Europese universiteit, wat vervolgens zou helpen bepalen of de GRE in de praktijk zou kunnen worden toegepast (tegenwoordig is dit meestal niet het geval). De GRE is een gestandaardiseerde test die uit drie schalen bestaat: verbaal redeneren, kwantitatief redeneren en analytisch schrijven. Het voordeel van gestandaardiseerd testen is dat het helpt bij het vergelijken van aanvraagdossiers omdat de GRE-scores op dezelfde schaal liggen voor kandidaten van alle soorten (nationale en educatieve) achtergronden.

Om te bepalen of de GRE potentieel nuttig is voor toelating tot Engelstalige onderzoeksmasterprogramma's in Europa, hebben we gegevens verzameld van 167 startende masterstudenten: zij namen de GRE General af en gaven ons informatie over

hun eerdere studiesucces, opleidingsachtergrond, en demografische kenmerken. Om deze gegevens te analyseren, hebben we hiërarchische regressieanalyses uitgevoerd. We toonden aan dat alle GRE-schalen GGPA voorspelden. Individuele GRE-schalen voorspelden het stagecijfer en de beoordelingen door supervisors van de onderzoek-gerelateerde vaardigheden van studenten, hun praktische vaardigheden en de structuur en stijl van hun onderzoeksrapporten. Geen van de GRE-schalen voorspelden de beoordeling door supervisors van de inhoud van onderzoeksrapporten van studenten. De geïdentificeerde relaties blijven staan nadat rekening is gehouden met sociaaleconomische status.

We concludeerden dat de GRE een redelijke voorspeller is van graduate studiesucces bij Engelstalige STEM research masters. We beschrijven ook de voor- en nadelen van de GRE-toepassing bij selectieve toelatingen (we beschreven ook kort de wettelijke beperkingen).

Hoofdstuk 6. Dit hoofdstuk geeft een samenvatting van de belangrijkste bevindingen van dit proefschrift. Hier worden de theoretische en praktische bijdragen van dit onderzoek en de beperkingen beschreven. De toelatingsbeoefenaars worden verwezen naar de sectie met de titel "Practical Contributions/ Implications for Applicants & Practitioners" van dit hoofdstuk om inzichten, richtlijnen en mogelijk inspiratie voor hun dagelijkse werk te krijgen.

De rest van het hoofdstuk is gewijd aan de beschrijving van een aantal belangrijke thema's die niet in detail aan bod kwamen in dit proefschrift (omdat ze buiten de reikwijdte vielen). Deze thema's vormen echter een belangrijk discours binnen onderzoek en praktijk van selectieve toelating. De thema's zijn als volgt: (1) verdienste en diversiteit ("merit and diversity"); (2) de effecten van selectie op de middelbare school op de diversiteit van leerlingen in het hoger onderwijs; (3) de voor- en nadelen van zelfselectie; (4) innovatieve benaderingen van studentenselectie zoals modulaire benadering, curriculum sampling of proefstudie, signaaldetectie theorie ("signal detection theory"), beperkte optimalisatie ("constrained optimization"); (5) alternatief model van selectieve toelatingen: combinatie van selectieve toelatingen met loterij (in de ex aequo -groep ["loterij in het midden"] of een andere versie zoals "loterij van de gekwalificeerden"). De bespreking van deze brede thema's gecombineerd met de conclusies uit onze eigen studies leiden ons tot de conclusie dat het model van pure selectieve toelatingen in de huidige vorm misschien niet het beste model is om de doelen van eerlijkheid, objectiviteit en transparantie te dienen bij toelatingen: Toelatingsbeoefenaars moeten open blijven staan voor alternatieve manieren van toelating. Onderzoekers moeten doorgaan met het in twijfel trekken van elk van die modellen op de best mogelijke manier om het uiteindelijke doel van gelijkheid te bereiken en om de groeiende, diverse studentengemeenschap zo goed mogelijk te dienen.

Acknowledgements

How could I possibly express all the gratitude to you, dear colleagues and friends, in just a few pages of text? Even though it is not really possible, I hope that reading these few lines will give you a hint of my thoughts and feelings toward you across these years.

I felt your engagement as well as care and attention from the very start of our interactions: I think your engagement is what made my academic path across the last years and this PhD project a success. Who could have thought that a period of writing a dissertation, described by some others as emotionally difficult and stressful, would be the total opposite for me. For me personally, it was a journey of my dreams, and I am so grateful to all of you who made it possible and made it that pleasant.

First of all, I would like to thank ***my three amazing supervisors***, who stood next to me all these years. I felt your enormous, constant, and unconditional support. Thank you for your trust in me. I was so lucky I got a chance to work with the three of you and to learn from you.

Harold, thank you for your enthusiasm, bright ideas, accessibility, attention, and empathy. I know I can always be open to you about everything. This opportunity to share my thoughts, worries, or hopes with you, to be heard and understood was one of those secret ingredients that helped me to become more confident in myself as a researcher and as a person. Thank you for that feeling of freedom that you gave me, as well as that thoughtful advice that you provided at any moment I needed it. You taught me that there is a variety of possibilities, many things are achievable, and the world is open and listening. Following your kind guidance, I have been learning to enjoy the path, its beauty, and the learning experiences I get on my journey to my own Ithaca, and I am so grateful to you for that.

Marijk, that "Dare to Cross Over" event when I met you was the start of an important path for me, both professionally and personally. This path has opened my mind, enriched my views, challenged me to try my best, and ultimately changed so much in my life in a positive direction. Thank you for showing me by your own example how to stand for what I believe in and for your confidence in me. Across the years that we know each other, you regularly suggested me to step out of my comfort zone to try to do something unknown. After daring to do so, I always felt how much I grew as a person. I am so grateful to you for that. Your professionalism and high standards that you applied to this project, me, and yourself have been an inspiration for me and I will try to adhere to them no matter where my path will lead me.

Gönül, thank you for being there for me during all the ups and downs, for always finding time for me, and repeating constantly that I can reach you any moment. I felt I could always turn to you and be open about my worries. Your involvement in each study and your organizational talent were so helpful in moving challenging projects forward.

I want to especially thank you for your constant positive attitude, which energized and motivated me even at those moments, when I felt unenthusiastic or frustrated. Your sense of what your mission is and your commitment to it have been an example for me all these years, and I try to search for those internal landmarks in myself. You gave me freedom and encouraged me to make my own decisions, which gave me that very special feeling of empowerment, and it means so much for me, thank you.

My dear paranymphs, **Farah and Lieselotte**, since the moment I started imagining my defense, I knew that the two people whom I want to be right behind me during those very special 45 minutes of my life, are the two of you. I guess that's because all these years that we know each other, I felt such a positive connection with the two of you, and I knew I can totally trust you just to be there for me in the most "spannend" moment of my academic career. Thank you for all those instants of joy and laugh, as well as for those moments of consideration and attention that you gave me during this last, most responsible period before my defense. They made it so special, and I cannot express enough how happy I am that these few months also made the three of us grow closer, I value this a lot.

I would also like to thank all ***BMS and BMW colleagues*** who were always so attentive, helpful, and fun. I know for sure that this project would not have been a success without all of you. You all were my best teachers of the Dutch language. You were so attentive at various moments, and created the amazing atmosphere where I, in spite of my international background, could feel myself safe, included, and accepted.

Pauline, I have always been amazed by how hard-working, goal-oriented, and quick you are. I am glad that our work on the projects here and there gave me a chance to get to know you better as a person: Sometimes I caught myself on the thought that "hey, just stop overthinking, look at Pauline, she would already be through with it, and would not allow any small problems or opinions of others to stop her project or spoil her mood for long. Just do the same and it all will be fine". Thank you for being like that and taking me along your admissions projects.

Rianne, how nice it was to have you as a more experienced educational researcher nearby all these years. I remember how you took me for a drive to VU Amsterdam for the defense of Anouk Wouters just a few days after I joined the BMS team. We barely knew each other, but that one drive was enough for me to feel as if we had been colleagues for years. Time was passing by, but your warmth, your welcoming and helpful attitude have stayed the same, and I really value them.

Marjan, your unconditional support and advice on any question (of which I had a lot!) was so important for me that it is hard to overestimate. Those small things, which drove me crazy, always found an easy and quick solution once I turned to you

for an advice. And I cannot tell you enough how much I appreciate that you always took that role of integrating me into a Dutch-speaking team and how important it was for me, especially at the start. I can remember many examples, but most importantly, your constant attention to me as a person and to any difficulties that my international background could cause, meant a lot to me.

Michele, I felt I could turn to you with anything, and I would find a thoughtful advice and a helping hand. You have always been there for your colleagues, including myself. Your personal story and your passion for co-creation in education practice are so wonderful that they continue to inspire me.

Renske, meeting you was so refreshing for my attitude toward educational research. Your passion for doing research, for thinking about theory and methodology, your professionalism, determination, and openness about any issues inspired me a lot and I am very grateful to you for that.

Shirrinka, you have been an example for me on how to be in control of your work and not only that. I really appreciate your help with all kinds of organizational issues, which were essential to move this PhD project further, your engagement, and most of all, your professionalism and personal example.

Christine, once I started working with you at the end of my project, I realized that I got a lucky ticket. Your dedication, your motivation to reach the best possible outcome, your ability to relate to and to focus on the needs of another person and their project astonishes me. You taught me a lot, and the rules of English grammar are the least of it. Thank you for taking your work with me so seriously, for believing in the quality of my studies, and for teaching me to make my voice sound confident in my academic writing.

Elena, it was so important for me to know that you were there, not as an always busy administrator, but as a person, who cares: about her colleagues and about their work, including me and my project. I do not think you even realize which role you played in the success of some of my studies. Likewise, **Yolande and Mariska**, I am so grateful to you and your whole team that all the communication about the studies included in this dissertation went smoothly and that you were always ready to help in spite of all the urgent administrative school matters you had.

Nivard, I remember how Gönül advised me to contact you to talk about working with Osiris data, saying that you will probably be the one with whom I will work most closely in the team. It turned out to be so true: you were not only one of my closest colleagues, but also someone to whom I could turn for an advice in any unclear or uncertain moment. You took our collaboration very seriously and it made me feel that we are doing something important together. Thank you for those brainstorms, for your humor and bright outlook on life, which always raised my mood.

Astrid, I felt so special being your direct desk neighbor in the 4.04 office before we started flexworking. Your organized manner of working, your respect toward all your colleagues, including myself, your attention to details including sending those nice

birthday emails to me every year, made me always feel welcome in our big team. Thank you for creating this atmosphere.

Kirsten B. and **Sanne,** thank you so much for always being ready to answer all my questions about students' research projects, program coordinators, and anything else. Thanks for your positive attitude and the cozy atmosphere you created since the moment I started the project.

Gisela, I felt that you could relate to my (personal) experiences. To say the least, you know very well how it feels when one's loved ones (such the parents) live on a huge distance. A certain similarity of our backgrounds, and other small, invisible moments of our interactions made it such a great experience. I knew I could always turn to you. Thank you for that.

John, thank you for your calm and kindhearted attitude, I enjoyed our conversations about research and life. You respect toward research and your colleagues, as well as your thought and time that you invest in each project have always been very encouraging for me.

Cecily, I really appreciate your help with coding those hundreds of articles for the review study. You were the best companion to do this big amount of work.

Simone, jij bent degene die mij constant uitdaagt om mijn Nederlands te verbeteren. Ik herinner me hoe je eerste vraag aan mij was wat mijn verjaardag is. Je schreef het op een papier dat je in 4.04 inleverde en vanaf dat moment voelde ik me onderdeel van een team. Bedankt voor je goede humeur, je glimlach, en niet te vergeten, je reflectieve nieuwjaarsmail.

Krijn, Monique, Kirsten K., Maarke, Marc, Ellen, Juliette, Tania, Heleen P., Els, Heleen R., Anne-Petra, Spiros, Eliane, Isabella, Dimitra, and Ouafa, it was amazing to work with you in one big room! Thank you so much for all those moments when you offered a helping hand, made a joke, or shared a personal story. Thanks to these our conversations in and outside our office, which made me feel included into our big team and I knew that I could approach you whenever I feel like that. I have many small stories in my memory—too many to mention here, but you should know that I feel myself very lucky to have the colleagues like you.

I would also like to take a moment to thank ***the colleagues outside of my team***.

I would like to express my gratitude to **all GSLS program coordinators.** You allowed me to get an intimate look at the selective admissions of your programs, contrasted my research ideas against the practical issues you encountered in your daily work, you asked difficult questions, challenged me and by doing so, you helped me to make this project a success. I would like to express a special thanks to **Kristin, Bertrand, and Joost** for participating in the pilots that Pauline, Shirrinka, and I initiated. Even though those pilots are not part of this dissertation, they are invisibly included in some considerations, expressed in this book.

Quirine, Anton, and Ivana, I am so glad I had an opportunity to work with you. Thanks to the three of you with your interest in my research, your thoughtful comments here and there, and most importantly, your personal attitude. I value our good relationships a lot.

Linda, I remember well how amazed I was after reading your first published paper and later, your dissertation. Even before getting to know you personally, I knew that this work can be done only by someone who not only thinks very carefully about all aspects of admissions research, but also cares about students and the role that the admissions decisions play in their life. I am grateful to you for your positive personal and organized professional attitude and showing an example of how to proceed making good quality research in spite of all the workload.

Judith, how glad I am that our good relationship went on beyond those years when I was your master's student. Thank you for just being there: in the back of my head, I always knew I could drop you an email and come by your office. I love our rare, but warm coffee moments, when I can talk to you about anything.

Bart, Dennis, and Hans, I could not imagine that you, those admissions officers (who seem invisible to applicants and who, I guess, processed my application for the DaSCA program and the Utrecht Excellence Scholarship), will once become my colleagues and I will discuss the challenges of admissions with you. I am very thankful for all your engagement and help across the years we know each other.

Guus, bedankt dat je ons, je indirecte collega's, hielp met jouw professionele advies. Ik waardeer echt alle tijd die je me hebt bespaard toen je Nivard en mij hielp om de technische details van gegevensquery's te achterhalen.

Claudia, I really appreciate your help with collecting data for the GRE study.

Hanneke, thank you so much for your help with and advice on the UU data

Yaran, thank you for proofreading the Dutch version of the summary to this dissertation.

And **all those colleagues** whom I forgot to mention: Thank you for being helpful across all these years of my project.

Not least, I would like to thank ***all the applicants and GSLS, GSNS, and GSGS students that participated in the studies*** conducted in this dissertation. This research is done first of all with keeping the interests of applicants of different backgrounds in mind, and I hope that this attitude is visible enough through the pages of this dissertation.

I would also like to take a moment to thank ***my good friends***.

Marina and Dasha, my two guiding stars. I might have told you this before, but I will repeat it here in writing: You were those two people whose example made me leave the

known and make that one very crucial step in my life— to apply for a graduate program in the West. Your amazing (though also challenging) French and American journeys made me believe that everything is possible. I told to myself that if you could manage to go through those first most difficult years of immigration, I will be able to manage it as well, despite all the difficulties and uncertainty. I am not even sure that you are fully aware of the roles that you both, your determination, your values, your ups and downs, played in my own life and ultimately in the fact that I am writing these words to you in my dissertation book. My friendship with both of you is the best outcome I got from five years at the MSU, and I am grateful to that period for just having the two of you in my life.

Yan, who could have imagined that a random decision of a SSH administrator to assign us rooms in one shared apartment will turn out to become such a good friendship. I have always been amazed how wise and thoughtful your kind advice was, and I know that I can always rely on you no matter what happens. I am so glad that we had been walking this immigration path together (even though in different ways, of course), and that I do not need to send this book to you to China: I am just enjoying having such a good friend nearby, in the country that became home for both of us.

Jessie, how happy I am about that question on that September day of 2014, our friendship is the best answer to it. Your kindness, intelligence, and responsiveness are those features which I admire, and I thank you for being there all these years for me. I am so happy that the life granted me this opportunity to get to know you, as well as to stay in close contact years after we graduated from DaSCA.

Liza, Dasha, and Sveta, how nice it is to have you as my friends. I am so happy we met and your presence in my life, our long talks when we see each other make my life joyful and I am grateful to you for that.

Тамара, спасибо вам за то, что вы есть в моей жизни. Я всегда чувствовала вашу безусловную поддержку[1].

Finally, I would like to say a few words to ***my loved ones***.

Matthias, my Love, what an amazing present I got from life when I met you. I started my PhD project just a few months after we met, and you witnessed it all. You stood next to me at every moment: no matter how challenging or joyful those moments were. I am so lucky you were always there for me, as a partner and as an outstanding, bright researcher. Some of your methodological ideas, statistical advice, as well as all other kinds of suggestions are integrated in this book. They were born during many of our trips, skype calls, late evening walks, relaxed cooking, or eating your amazing pasta Bolognese. Once you asked me jokingly what would I do without you, and I am very

1 If you are curious to know what it says, it translates something like: "Tamara, thank you for being and being in my life. I have always felt your unconditional support"

seriously answering it now that I do not know what I would do, because I barely can imagine my life right now without you by my side. You are the reason why I can call the last six years of my life not only successful, but also very happy. Ich liebe Dich, mein Schatz.

I would also like to thank my German family. **Ute, Karlheinz, Katharina**, thank you for making me feel at home in Ingelheim, for your care and support, for those family trips, and long warm, cozy evenings with a big family that came into my life together with you. I am very much looking forward to even more of them in the years to come.

Дорогие мои, **любимые мама и папа**, все, чего я достигла в жизни, включая написание этой диссертации, во многом было определено вами: вашей верой в меня, вашим вкладом в моей образование, начиная с самого моего детства, и вашей любовью. Вы всегда ставили мое образование в приоритет и отказывали себе во многом, ради того, чтобы у меня в жизни появились перспективы. Мне это казалось таким естественным в детстве, но только сейчас, заканчивая написание своей докторской диссертации и имея в своей жизни то, о чем я даже мечтать не могла, я начинаю по-настоящему осознавать, что на самом деле вы для меня сделали. Спасибо вам обоим за это, за вашу заботу, за ваше безусловное принятие и любовь. Я очень сильно вас люблю[2].

Thank you all for reading so far. This was a wonderful journey, in which the path itself turned out to be the best imaginable learning experience I ever had. It would have never been possible without all of you. I hope to see all of you again to tell you already in person how grateful I am for all your support and care.

September 2022.
Utrecht.

2 My dear, beloved mom and dad, everything that I have achieved in life, including this dissertation, was largely determined by you: your faith in me, your contribution to my education since my childhood, and your unconditional love. You have always made my education a priority and limited your own needs in many ways, so that I would have prospects in my life. It seemed so natural to me as a child, but only now, as I finish my doctoral dissertation and have such a life that I could not even dream of, I begin to truly realize what you have actually done for me. Thank you both for this, for your care, for your unconditional acceptance and love. I love you very much.

Curriculum vitae

Anastasia Kurysheva was born on August 31, 1988 in Tashkent, Uzbekistan (a.k.a. Uzbek Soviet Socialist Republic before 1991). After completing her secondary education in Gymnasium №50 in Tashkent, she studied psychology at Moscow State University, obtained a merit scholarship, and graduated as a clinical psychologist and teacher of psychology. She worked in clinical settings for three years as an educational and neuro-psychologist, specializing in neuro-psychological assessments and support of children and adolescents with cognitive and emotinal impairments. Trying to aid those kids in becoming more successful, she became interested in research on cognitive determinants of school study success. She applied to two-year research master's program at Utrecht Univerisity in "Development and Socialization of Children and Adolescents", which she completed in 2016.

Motivated by her research interest in cognitive determinants of study success, she turned to practical and research work in the field of Higher Education. She worked in a position, in which she was assisting the Dean of Graduate Studies in the development of the Learning Community on Admissions and a number of other projects. She later obtained a PhD position at the Educational Center University Medical Center Utrecht with the focus on determinants of study success at a graduate level. She also taught statistical courses. Combining her expertise in educational studies, psychological assessment, and statistics, Anastasia has been researching admissions criteria of research masters' programmes, with a specific focus on their validity, acceptability by stakeholders, and transparency. In applied settings, she is working on two projects of the Utrecht University Graduate School of Life Sciences: one of them pilots standardized recommendation letters and another one pilots forms of admissions, alternative to purely selective admissions. Next to her academic work, she is a co-founder of Radio Life Sciences—a podcast series with educators, students, researchers, and policy makers within the fields of Higher Education and Life Sciences.

Publications

In this dissertation:

Kurysheva, A., Koning, N., Fox, C. M., van Rijen, H. V., & Dilaver, G. (2022). Once the best student always the best student? Predicting graduate study success, using undergraduate academic indicators. Evidence from research masters' programs in the Netherlands. *International Journal of Selection and Assessment*, 1-17. https://onlinelibrary.wiley.com/doi/10.1111/ijsa.12397

Kurysheva, A., van Rijen, HVM. & Dilaver, G. (2019). How do admission committees select? Do applicants know how they select? Selection Criteria and Transparency at a Dutch University. *Tertiary Education and Management*, 25, 367–388. https://link.springer.com/article/10.1007/s11233-019-09050-z

Not in this dissertation:

Kurysheva A., van Ooijen-van der Linden L., van der Smagt M., & Dilaver G. (in press). The added value of Signal Detection Theory as a method in evidence-informed decision-making in Higher Education: A demonstration. *Frontiers of Education. Section Assessment, Testing, and Applied Measurement*. doi: 10.3389/feduc.2022.906611

Torotcoi, S., Gologan, D. & Kurysheva, A. (2020). What works for underrepresented groups? Identifying effective measures for enhancing access to higher education in Europe. In A. Curaj, L. Deca, & R. Pricopie (Eds.), *European Higher Education Area: Challenges for a new decade* (pp. 177-196). Cham: Springer. https://link.springer.com/chapter/10.1007/978-3-030-56316-5_13

Summary of PhD training

Type of training	*Workload (EC)*
Courses	
Understanding internationalization of Higher Education in a changing global context.	1.5
Science and politics: Exploring the relations between academic research, higher education, and science policy.	5
Roles of universities in European integration	5
Data science: multiple imputation in practice	1.5
Advanced multilevel analysis	1
Analytic storytelling	0.7
Giving effective oral presentations	0.6
Scientific artwork with Adobe Photoshop and Illustrator	0.6
Adobe InDesign essentials	0.6
Responsible research through supervision, mentoring and working together	0.5
Intercultural communication	0.3
Personal development and competences during your PhD	0.1
Research planning and time management	0.5
Writing a scientific paper	2
Stress management	0.3
Modeling multilevel data with small sample sizes	
Dynamic multilevel modeling with Mpul8	
Systematically searching the literature	
Funding proposals	
Fit for the Future research meetings	
Total workload of courses	20.2

Conferences, symposia, and seminars

Challenge-driven, Accessible, Research-based, Mobile European University (CHARM-EU) Inclusion conference. (16 September, 2022; *online*). Member of a panel "Sharing experiences & networking: Making policy personal"
Junior Researchers of EARLI (JURY) conference (*Porto, Portugal*; 18-22 July, 2022). "Predicting Graduate Study Success Using Undergraduate Academic Indicators". Award for the best round-table
European Universities- Critical Futures [A workshop/PhD course] (14 December, 2020; *online*). A presentation "Selective Graduate Admissions to the STEM Master's Programmes"
VSNU Conferentie Internationalisering: Grenzeloos onderwijs als motor van de regio (*Venlo, the Netherlands*, 16 January, 2020). Participation
Consortium of Higher Education Researchers (CHER) conference *(Kassel, Germany;* 28-30 August, 2019). "Potential and Unique Incremental Value of Multilevel Modelling and Signal Detection Theory Methodologies in Analysis of Determinants of Graduate Academic Achievements and Failure at Research Master's programmes"
14th Conference of the European Sociological Association (ESA) *(Manchester, England;* 20-23 August, 2019). "Developing Diversity Sensitive, Evidence-Based Student Selection Toolkit"
European First Year Experience (EFYE) conference *(Cork, Ireland;* 17-19 June, 2019*)* "Determinants of successful transition from undergraduate studies to graduate studies: literature review and study results*"*
European University Association (EUA) conference *(Ghent, Belgium;* October, 2019). Participation.
Onderwijs Parade (*Utrecht, the Netherlands*, 2019). A workshop "Selective admission process and its impact on inclusion"
Consortium of Higher Education Researchers (CHER) *(Moscow, Russia;* 30-31 August, 2018). "Selective Admission Criteria for Life Sciences Master's Programs"
European First Year Experience (EFYE) conference *(Utrecht, the Netherlands;* 25-27 June, 2018). "The PhD Project Overview on Graduate Selective Admissions and the Survey Study on Selective Admission Criteria and their Transparency"
Association of the Dutch Universities (VSNU) conference *(Amsterdam, the Netherlands;* 21 May, 2018*).* Workshop:"Unbiased admissions: achieving equality in higher education"
The International Medical Educators Exchange (IMEX) Program (*Utrecht, The Netherlands*; 23-27 October, 2017). "Graduate Admissions & Student Study Success at Life Sciences Masters' Programs"
Multilevel Conference (*Utrecht, the Netherlands*; 12-13April, 2017). Participation.
MasterMind Europe conference *(Amsterdam, the Netherlands;* 5-7 July, 2017). "Validating the Admissions Process at Utrecht University Graduate School of Life Sciences"

A reviewer for the academic journals

Computers & Education
Tertiary Education and Management
Advances in Health Sciences Education

Other activities

Participation in a working research group "European Universities-Critical Futures Refugee Access to HE"
Facilitation "Learning community on admissions"
Participation in a working group of policy makers "Alternative admissions"
Facilitation of organization of a seminar "The graduate student experience at Utrecht University in international comparative perspective" (9 September 2019, Utrecht)
Participation in the Graduate Student Think Tank at Utrecht University 2016-2020 (co-chair)
Organization of a scholarship event 2017 for students of Utrecht University
A co-founder, an interviewer, and a facilitator at the Radio Life Sciences - a series of podcasts