RAISING THE BAR

Higher education students' sensitivity to the assessment policy

Rob Kickert

Raising the bar:
Higher education students' sensitivity to the assessment policy

Rob Kickert

Erasmus Universiteit Rotterdam

Raising the Bar: Higher education students' sensitivity to the assessment policy

Cover design:	Jelle Ris
Layout:	Marilou Maes, persoonlijkproefschrift.nl
Printing:	Ridderprint | www.ridderprint.nl
ISBN:	978-94-6416-142-7

Raising the Bar:
Higher education students' sensitivity to the assessment policy

De lat hoger leggen:
De sensitiviteit van studenten voor het examensysteem in het hoger onderwijs

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. R.C.M.E. Engels

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

vrijdag 4 december 2020 om 11.30 uur

door

Robert Kickert
geboren te Heemstede

Erasmus University Rotterdam

Promotiecommissie

Promotoren:	Prof.dr. L.R. Arends
	Prof.dr. P. Prinzie
Overige leden:	Prof.dr. A.A.C.M. Smeets
	Prof.dr. C.P.M. Van der Vleuten
	Prof.dr. J. Cohen-Schotanus
Copromotoren:	Dr. M. Meeuwisse
	Dr. K.M. Stegers-Jager

Ik zeg ja maar jij zegt nee
Ik zeg 'ga' maar jij die bleef
Ik ga staan maar jij beweegt
niet,
Altijd ben je daar,
jazeker

En waak je over mij,
als bange dagen over hoop
Als wrange vragen
maak jij je van mij meester

En maakt dat ik wil weten
Dus stel ik slechts de vragen
Wie niet waagt die kan niets weten
Niet vragen naar problemen,

is vragen om problemen
Dus ga ik maar voor zeker
En zaai ik je
In alle lagen van mijn leven

Want ik ben bang van mijn angst
Heb geluk met mijn geluk
En treur om mijn verdriet
Maar twijfel niet,

aan mijn twijfel.

Contents

CHAPTER 1

General Introduction

"If you can handle me, you can handle the world"
- Patrick Kickert

This quote was my father's motto for raising my two older brothers and me. He implemented this motto into practice by setting very challenging standards. On the one hand, his rewards for meeting these standards were even more unpredictable than life itself; we never knew what to expect. On the other hand, his punishments were completely predictable; we always knew what to expect. And as my father made sure we really wanted to avoid these punishments, the stakes were high. Through his high standards and stakes, my father was always motivating us to improve ourselves; he made sure that we not only set high goals, but also showed perseverance in attaining those goals. Moreover, he helped us to develop our self-regulation; he made us proactive learners, by constantly making us reflect on our own feelings and behaviours. I must admit that as a child, my father's challenging childrearing style was not always easy, nor always fun. Fortunately, my dad made wise choices concerning which behaviour he would reward or punish. Hence, I really do feel very well equipped to handle the world now. In hindsight, my dad's challenges were a didactical act of love.

Through the implementation of his motto, my dad taught me a basic principle of education: you learn a lot by being challenged. In fact, higher education institutions essentially have the same motto as my father, albeit more implicit: if you can handle this curriculum, you can handle the world. However, there is an important difference between my father and higher education as well: my father's love was unconditional, whereas higher education institutions require students to meet the demands of the assessment policies in order to avoid academic dismissal and progress academically.

The aim of this dissertation was to examine whether changes to the assessment policies are related to student learning. More specifically, the first aim was to investigate whether academic progress and academic performance (i.e. grades) are associated with changes to assessment policies. The second aim was to elucidate why performance differs under different assessment policies. We used motivation and self-regulation as a conceptual framework, as these concepts are two of the most important predictors of academic performance (Richardson et al., 2012). However, to the best of our knowledge, students' motivation and self-regulation under different assessment policies have received scarce attention. Thus, we investigated mean level differences in motivation and self-regulation, as well as differences in the relations of motivation and self-regulation with performance under different policies. Our third aim

was to elucidate whether all students are sensitive to the assessment policy in a similar fashion. Therefore, we examined whether different types of students exist regarding motivation in the first year. Our fourth and final aim was to explain why students' motivation is sensitive to assessments, and what the implications of this sensitivity are. Therefore, we developed a theoretical perspective on student motivation in higher education.

The different studies of this dissertation were sparked by changes made to the assessment policy at all faculties of Erasmus University Rotterdam (EUR), which created a rare natural quasi-experiment. These policy changes were made to accelerate academic progress, as swift academic progress saves time, money and energy for society, higher education institutions, and students. Therefore, accelerating academic progress is an important aim, both in Europe and the United States (Attewell et al., 2011; Vossensteyn et al., 2015). Before we discuss our theoretical framework, the conceptual model of this dissertation, and the aims of the different chapters of this dissertation, we will describe why and how EUR changed the assessment policy.

The Context for a New Assessment Policy

Traditionally, the Dutch government has acknowledged the importance of giving students ample time to learn at their own pace during higher education. For instance, in 1986 all Dutch higher education students got a six-year basic scholarship (Dutch: *basisbeurs*), whereas most programmes were only four-year programmes (Studiefinanciering door de jaren heen, 2012). Thereby, students were explicitly allowed to take more time to finish their course programmes than was essentially required. However, this six-year scholarship system soon turned out to be financially untenable (Strikkers, 2015). Consequently, in 1991 the basic scholarship for each four-year programme was reduced to five years, and to four years in 1996. In other words, students were now expected to finish the course programme in time, or to pay for their delays themselves.

Three additional measures were taken by the government to ensure the financial system would be tenable in the long run. Firstly, an academic dismissal policy called a Binding Study Advice (BSA; Dutch: *Bindend StudieAdvies*) was introduced in 1993 (for a description see Arnold, 2015). The BSA entails that Dutch higher education institutions can disallow first-year students to reregister for their course programme for the following three years, if the students' number of attained first-year credits is below a threshold determined by the university. Most higher education institutions set this threshold between 34 and 45 out of 60 credits. The purpose of the BSA was to prevent students from lingering unsuccessfully in their higher education programme

for too long, and instead switch to a more suitable programme. Thus, the BSA has both a selective and a referential function (Arnold, 2015). Both the selection of potentially successful students and the referral to a more suitable course programme should save time and money for society, higher education institutions and students.

Secondly, in 2012 the government made so-called performance agreements with higher education institutions (Boer et al., 2015). As a consequence, part of the institutions' funding was made contingent upon students' academic progress. For instance, EUR made the performance agreement to raise the four-year Bachelor's graduation rate from 69% to 75% of all students who started the second academic year (Reviewcommissie Hoger Onderwijs en Onderzoek, 2016).

Thirdly, the government lowered the amount of funding per student. Whereas the number of enrolling students increased drastically between the years 2000 and 2010, the government funding has not increased accordingly (VSNU, 2012). Hence, higher education institutions needed to be more economically efficient. In conclusion, the government has made satisfactory academic progress rates a key condition for universities' healthy financial status. Given these financial incentives to optimise academic progress, in 2011 EUR started to implement *Nominal is Normal,* an adapted version of the BSA.

A New Assessment Policy: Nominal is Normal

Originally, the main goal of the BSA was not to accelerate academic progress, but to improve selection and referral (Arnold, 2015). And indeed, studies on the traditional BSA of 34 to 45 out of 60 first-year credits indicated no differences in obtained credits or first year completion rates (De Koning et al., 2014; Eijsvogels et al., 2015; Stegers-Jager et al., 2011). When considering all enrolling students, Bachelor's graduation rates also did not change due to the introduction of the BSA (Arnold, 2015). The selection in or after the first academic year did seem to change: first-year student dropout increased by 5.8 - 7.5%, and four-year Bachelor's graduation rates for the students who re-enrolled in the second year were 3.3 – 9% higher (Arnold, 2015; Sneyers & De Witte, 2015). Arnold (2015) expressed his concerns regarding the referential function of the BSA, as "the BSA does not prevent students from languishing in higher education" (p. 1081).

In an attempt to accelerate students' academic progress and improve the selective function of the BSA, EUR introduced an assessment policy called '*Nominal is Normal*' (N=N; Vermeulen et al., 2012). Under N=N, all faculties changed the *stakes* of the assessment policy. In addition, most faculties also changed the *performance standard* and the *resit standard*.

The *stakes* for assessments concern the consequences of failing assessments. These consequences are a result of the timeframe in which students are required to obtain all first-year credits, and thereby evade academic dismissal. Under the old policy, first-year students were required to obtain 40 out of 60 first-year credits within one year, and all 60 credits after two years (Vermeulen et al., 2012). However, the 40-credit minimum within one year turned out to have the adverse impact of becoming a target for many students, instead of a minimum (Arnold, 2012; Stegers-Jager et al., 2011; Vermeulen et al., 2012). Many students seemed to lower their efforts once the threshold of 40 credits had been reached. In addition, during the second year, the non-completed first-year courses may compete with the second-year courses (Stegers-Jager et al., 2011). Therefore, under N=N students are required to obtain all 60 credits within one year, and thus the stakes are higher than under the old policy.

The *performance standard* concerns the passing grade needed to obtain credits for a single course (i.e. subject). In the Netherlands, this performance standard is usually conjunctive, which entails that each individual single course needs to be passed (Chester, 2003; Yocarini et al., 2018). In contrast, a compensatory performance standard allows some form of compensation between grades for separate single courses. A compensatory standard should result in more reliable decisions about students' progress or dismissal within a cluster of correlated courses (Yocarini et al., 2018). Therefore, although the specific changes to the performance standards were different per faculty, most faculties changed their performance standard under the new policy by making it compensatory.

The *resit standard* concerns the number of permitted resits. Lowering the number of resits was intended to reduce procrastination among students (Vermeulen et al., 2012). Consequently, compared to the old policy, under N=N the numbers of possible resits in the first academic year were lowered in many faculties. The changes to the stakes, performance standard and resit standard of the assessment policy at EUR provide a unique opportunity to investigate the consequences of these changes for student progress. Therefore, an important aim of this dissertation was to investigate whether progress indeed accelerated as intended, and if so, how this could be explained.

N=N and Academic Progress

Several research reports have shown that under N=N, student progress after one academic year is higher in the first N=N cohort than under the old policy (Baars et al., 2013; Vermeulen et al., 2012). Vermeulen and colleagues (2012) reported about a pilot study of N=N involving the social science course programmes of EUR: the proportion of students who obtained all first-year credits after one year was 21-39% higher under

N=N than under the old policy. In a study that included all but the medical students, Baars et al. (2013) concluded that on average the percentage of EUR students who completed the first year within one year increased from 35% to 59%. Thus, progress was faster under N=N than under the old policy.

In addition to comparing one-year progress, it is possible is to compare final progress, which means progress after one year under N=N versus progress after two years under the old policy. Vermeulen et al (2012) reported that in the social sciences, final progress is 5-9% higher under N=N than under the old policy. Baars et al. (2013) concluded that EUR-wide, again not including medical students, the proportion of students who obtained all first-year credits after one year under N=N, was comparable to the proportion of students who obtained all first-year credits after two years under the old policy. In sum, the differences regarding final progress under N=N versus the old policy were smaller, or absent.

As the reports on progress under N=N versus the old policy (Baars et al., 2013; Vermeulen et al., 2012) only included the first N=N cohorts and did not include medical students, the first aim of this dissertation was to further investigate differences in progress between the old and new assessment policies. Therefore, we compared academic progress under the old assessment policies versus the new N=N-policy in three large faculties at EUR, i.e. Business Administration, Medicine, and Psychology.

Explaining Differences in Academic Progress Between Assessment Policies

The abovementioned differences in progress do not seem to be explained by a selection effect before the start of the first academic year. The number of students enrolling at EUR and the market share of most educational programmes at EUR has only slightly increased after the introduction of N=N (Baars et al., 2013). Thus, it does not seem to be the case that students are scared off by the new policy. In addition, the composition of the enrolling student population was shown to be generally comparable in terms of gender, age and various pre-university education characteristics (Baars et al., 2015). The only observed significant differences were small: slightly lower percentages of both students with a non-western migration background and of students with a preparatory-university high school diploma (Dutch: VWO) were found under N=N. Thus, differences in inflow characteristics do not explain the observed differences in the proportions of students who obtained all first-year credits.

Therefore, we investigated two other possible mechanisms through which differences in assessment policies are likely to affect academic progress. Firstly, a different policy may cause students to perform differently, i.e. achieve different grades. Secondly, it

may be that the selection for progress in the BSA decision has changed: grades that would have sufficed to progress to the second year under the old policy, may not suffice under the performance standard of N=N, or vice versa. Consequently, different student groups will progress under the old policy than under N=N. Thus, N=N may make a different selection for progress than the old policy. Therefore, in addition to investigating possible differences in progress, we investigated whether these possible differences occurred because of different student performance and/or different selection for progress.

Explaining Performance Differences Between Assessment Policies

Next, in order to explain differences in student performance (i.e. grades) under different assessment policies, in this dissertation we used motivation and self-regulation as our conceptual framework. The rationale to do so was threefold. Firstly, motivation and self-regulation are two of the most important constructs used in the explanation of academic performance (Richardson et al., 2012; Schneider & Preckel, 2017). Secondly, compared to other important predictors of academic performance, such as personality (Poropat, 2009), high school grades (Sawyer, 2013), or socioeconomic status (Sirin, 2005), motivation and self-regulation are relatively more alterable, and thus more likely to be affected by assessment policies. Thirdly, there was a scarcity of available literature on differences in motivation and self-regulation after comparable changes to the assessment policy as in N=N. We were only able to find literature on differences in motivation (Knekta, 2017; Simzar et al., 2015; Sungur, 2007; Wolf & Smith, 1995) or self-regulation (Sundre & Kitsantas, 2004; Sungur, 2007) under assessments with no consequences (e.g. assessment does not count towards grade) versus assessments with consequences (e.g. assessment counts towards grade). However, in this dissertation we will compare assessments with consequences (e.g. two-year timeframe to obtain all first-year credits) to assessments with even higher consequences (e.g. one-year timeframe to obtain all first-year credits). In addition, to the best of our knowledge no studies have investigated differences in motivation and self-regulation under different performance standards or resit standards. Therefore, with this dissertation we aim to fill this gap in the literature concerning differences in motivation and self-regulation under different assessment policies.

Motivation

Schunk, Meece and Pintrich (2014) define motivation as 'the process whereby goal-directed activities are instigated and sustained' (p.5). The fact that motivation is a process instead of an outcome means that motivation is not directly observable (Schunk et al., 2014). This poses a major challenge for any motivational researcher. In the current dissertation, we have chosen self-report questionnaires as the tool to

measure motivation. As motivation cannot be observed directly, self-reports are better able to capture unobservable processes than alternatives such as direct observation or ratings by others (Schunk et al., 2014). In addition, questionnaires are relatively time efficient. As especially the lower-motivation students may not want to invest much time and effort, more time-consuming alternatives such as interviews or dialogues may result in a more biased sample. The efficiency also allows for the investigation of large groups of students.

Furthermore, the definition of motivation suggests that in order to find out what motivates a student, two things need to be established: the student's goals, and the student's perseverance in instigating and sustaining effort to attain that goal (Atkinson, 1957; Schunk et al., 2014). Old policy students could have various goals concerning the timeframe in which to obtain all 60 first-year credits. Conversely, under N=N, a specific and difficult goal is determined by the curriculum: obtaining all 60 credits within one year. Goal-setting research has consistently shown that specific, difficult goals lead to the best outcomes, as long as these goals are attainable (Locke & Latham, 2002).

In addition to various goals, students can differ in their perseverance in attaining those goals. On the one hand, previous research on BSAs (Arnold, 2015; Sneyers & De Witte, 2015) and on a comparable American policy called academic probation (Lindo et al., 2010), has shown that setting minimum standards promotes higher drop-out, which may indicate a cease of perseverance. On the other hand, the same investigations showed improved performance for those who remain in the programme, which may reflect increased perseverance.

Two constructs that recur in several motivational theories may be important in order to understand students' perseverance: beliefs about competence and value (Cook & Artino, 2016). Beliefs about competence concern the question 'Can I do it?', whereas value is about 'Do I want to do it?', or 'What will happen (good or bad) when I do it?' (Cook & Artino, 2016). A specific competence belief is self-efficacy, which in higher education is a student's judgement of the ability to learn and/or perform (Bandura, 1982; Richardson et al., 2012). Students with high self-efficacy are more likely to persevere and work hard in order to learn, than students who doubt their own ability (Schunk et al., 2014). A specific measurement of the value construct is students' task value, which indicates the extent to which a student finds the material interesting and worth learning (Credé & Phillips, 2011). As for self-efficacy, we expect high task value to be a force of perseverance in learning. Students' goals, beliefs about competence and value are the most important motivational predictors of academic performance (Richardson et al., 2012). Therefore, regarding motivation we empirically investigated

students' goals, beliefs about competence and value after changes to the assessment policy.

Self-Regulation

In addition to motivation, self-regulatory factors are important correlates of academic performance (Richardson et al., 2012; Sitzmann & Ely, 2011). Self-regulated learners are "metacognitively, motivationally, and behaviorally active participants in their own learning" (Zimmerman, 1986, p.308,). Thus, the first element of this definition is metacognition. Self-regulated learners are described as masters of their own learning, who monitor and adapt their learning process accordingly (Zimmerman, 2008).

The second element in this definition is motivation, which is thus considered a component of self-regulation by Zimmerman (1986). However, as other scholars consider motivation and self-regulation separate categories (Pintrich, 2003; Pintrich & De Groot, 1990; Richardson et al., 2012), the relationship between both constructs is somewhat diffuse. Therefore, in this dissertation we assume that monitoring and adapting motivation can be considered self-regulation. However, there are also motivational processes that are not self-regulated, but for instance externally regulated. In fact, this dissertation concerns an assessment policy in which students' goal is externally regulated, as N=N requires all students to attain all first-year credits within one year. Another definition of self-regulation, as "the self-directive processes and self-beliefs that enable learners to transform their mental abilities (…) into academic performance" (p. 166, Zimmerman, 2008) further supports the differentiation between self-regulated and non-self-regulated aspects of motivation. Therefore, in this dissertation we differentiated between motivation and self-regulation.

The third element of self-regulation is behaviour, which may be a mediator between motivation and academic performance (Credé & Phillips, 2011). Zimmerman (2008) explains that self-regulated behaviour is sometimes equated to resource management, which denotes students' capacity to manage available resources (Credé & Phillips, 2011). Examples of these resources are students' time and effort. Metacognitive, motivational and behavioural self-regulatory constructs are all significantly related with academic performance (Credé & Phillips, 2011; Richardson et al., 2012; Schneider & Preckel, 2017; Sitzmann & Ely, 2011).

Despite the importance of motivation and self-regulation for understanding student learning and performance, there is a lack of literature on differences in motivation and self-regulation under different stakes, performance standards or resit standards. Therefore, in this dissertation we will investigate differences in motivation and self-

regulation, and differences in the associations of motivation and self-regulation with performance, to explain performance differences between students under different assessment policies. We will now first present our general conceptual model for this dissertation, followed by an outline of the contents of the different chapters.

Conceptual Model

Figure 1 depicts the general conceptual model for this dissertation. As presented in this figure, students are in a curriculum, which consists of: 1) objectives, 2) instructional activities and materials, and 3) assessment (Anderson, 2002; Cohen, 1987). As the objectives determine the intended outcomes of learning, the assessment should be aligned with these objectives in order to be a good reflection of learning (path a). Next, the assessment may affect student motivation and self-regulation (path b), which in turn may influence academic performance (path c). This relationship of motivation and self-regulation with performance may be affected by assessment (path d). For instance, good self-regulation may not affect performance similarly on assessments of different quality. In addition to this indirect influence of assessment on performance, there can also be a direct effect (path e), as academic performance is not just a result of the student, but also of the assessment. For example, the difficulty of the assessment will affect the grades that students obtain. Therefore, academic performance is not completely placed 'in the student' in this model. Next, academic progress is affected by academic performance, as some grades suffice to pass whereas other grades are insufficient to pass (path f). This relationship between performance and progress is also affected by the assessment (path g). For instance, a grade that results in progress may no longer suffice after the performance standard of the assessment is changed. Paths a-g of the conceptual model represent the aims of Chapters 2-6 of this dissertation, which we will now present.

Dissertation Outline

In addition to the current introductory chapter 1, this dissertation contains four empirical studies (chapters 2-5), one theoretical paper (chapter 6), and a general discussion (chapter 7). Table 1 gives an overview of each chapter's aims, measures, samples, statistical analyses, and links with our conceptual model presented in the previous paragraph.

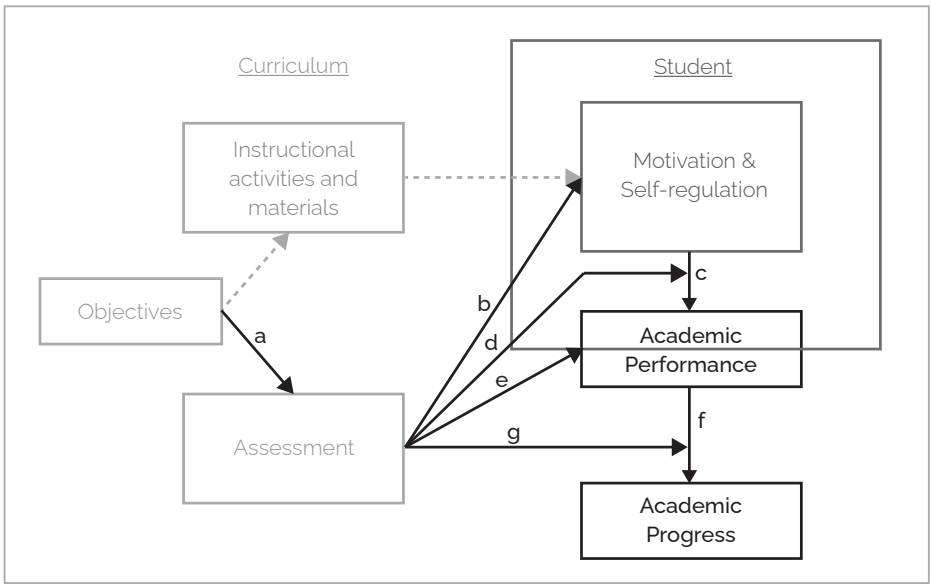


Figure 1. The general conceptual model for this dissertation. The black paths a-g represent the investigated associations in Chapters 2-6. Chapter 2 concerns paths e, f and g. Chapters 3 and 4 both concern paths b, c and d. Chapter 5 concerns paths b and c. Chapter 6 concerns paths a, b and c. The dotted paths are crucial for student learning but were not the topic of this dissertation.

The study in **chapter 2** concerns differences in first-year progress under the old and new (N=N) assessment policy. The three aims were to investigate the relationship between differences in assessment policies and differences in: 1) academic progress, 2) academic performance, and 3) selection for progress. We compared academic progress under the old and new assessment policies in three large faculties of EUR that made different changes to the policies: Business Administration changed the stakes; Medicine changed the stakes and performance standard; Psychology changed the stakes, performance standard and resit standard. In addition to progress, we compared students' GPA under the old and new policy in all three faculties. Finally, as Medicine and Psychology changed the performance standard, for both faculties we mimicked whether students would have progressed under the performance standard of the other policy. Thus, we mimicked old policy students' progress under the new policy performance standard, and vice versa. As well performing students should progress under different performance standards as well, we used this mimicked progress as another performance indicator, besides GPA. Additionally, we used the mimicked progress as an indication of differences in selection for progress between the old and new assessment policies.

In the study described in **chapter 3** we investigated motivation, self-regulation, participation and academic performance under two different assessment policies. Therefore, we used a previously validated structural model as our conceptual framework (Stegers-Jager et al., 2012). This structural model revealed positive associations between 'motivational beliefs' (i.e. motivation) and academic performance, that were mediated by 'learning strategies' (i.e. self-regulated learning) and 'participation in scheduled learning activities'. We first compared the average scores on motivation, self-regulation, participation in learning activities and academic performance of first-year medical students under the old and new assessment policies. Secondly, we examined whether the relations between motivation, self-regulation, participation in learning activities and academic performance were similar under the two assessment policies. To this end, we tested whether the structural model was invariant for students under both assessment policies. Students under the old and new assessment policy completed the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich et al., 1991) on motivation and self-regulation, and three additional items on participation (Stegers-Jager et al., 2012). Additionally, we operationalised academic performance as average official first-year grades, obtained from university records.

In the study in **chapter 4** we examined motivation, self-regulation and academic performance of two student groups who took the same statistics course, yet under different assessment policies: 3rd-year students of education and child studies (ECS) studied under an assessment policy with relatively higher stakes, a higher performance standard and a lower resit standard, compared with 2nd-year psychology students. Firstly, we compared academic performance of both groups of students, to see if we could replicate earlier findings on higher performance under more difficult assessment policies. Secondly, we compared both groups on the motivational and self-regulatory factors most strongly associated with academic performance (Richardson et al., 2012). Thirdly, we investigated whether the associations of these motivational and self-regulatory factors with academic performance are different under both policies. ECS and psychology students completed subsections of the MSLQ (Pintrich et al., 1991), as well as additional items on motivation. Both the official grades for the first attempt as well as resit grades were obtained from university records.

The aim of the study in **chapter 5** was to explore how students' motivation develops throughout the first year, and whether this development is the same for all students. To this end, we performed a latent class analysis on the two motivational factors most strongly associated with academic performance: students' grade goals and performance self-efficacy. Thereby, we explored how students shape their motivation around the performance standard, and whether different motivational classes of

students could be identified. To form the classes, we used data on grade goals and performance self-efficacy throughout the eight consecutive single courses of the first year for three samples of social science students. Next, we aimed to characterise and validate the classes by associating class membership with several student course evaluation items and official academic performance from university records.

In order to explain several findings from the studies in chapters 2-5, in **chapter 6** we present a theoretical perspective on student motivation in higher education. This perspective clarifies why it is adaptive for students to be sensitive to characteristics of assessments. Thereby, we also aimed to explain under which circumstances raising the stakes and standards may have negative consequences for student learning. Additionally, we give concrete suggestions for how these consequences can be remedied by our assessment practices.

Finally, in **chapter 7** we present a summary of chapters 2-6, and a discussion of the most important results and conclusions of this dissertation. This discussion entails an overview of strengths and limitations, as well as implications and directions for future research.

Table 1. Overview of chapters 2-6 of this dissertation.

Ch.	Aim(s)/RQs	Measures	Sample	Analyses	Paths
2	Are differences in assessment policies associated with differences in: 1. academic progress, 2. performance, & 3. selection for progress?	1. Academic Progress 2. a) GPA & b) Mimicked progress 3. Mimicked progress	First-year students: Business Administration (n = 2048); Medicine (n = 1630); Psychology (n = 1076)	1. Chi-squared tests 2. a. T-tests b. Chi-squared tests 3. McNemar tests	e, f & g
3	1. Do average scores on motivation, self-regulation, participation in learning activities and academic performance differ under the old and new assessment policies? 2. Are relations between motivation, self-regulation, participation in learning activities and academic performance similar under the two assessment policies?	1. & 2. Intrinsic goal orientation, task value, academic self-efficacy, elaboration, organisation, metacognition, time and study environment management, effort regulation (all from MSLQ), participation, average grades	First year medical students (n = 1177)	1. MANOVA & follow-up ANOVAs 2. Multi-group Structural Equation Modeling	b, c & d
4	1. Does academic performance of students under two different assessment policies differ? 2. Do students' motivation and self-regulation differ under both assessment policies? 3. Are the associations of motivation and self-regulation with academic performance different under both policies?	1. Initial and final (post-resit) grades, use of resits 2. & 3. academic self-efficacy, task value, effort regulation, time and study environment management, test anxiety (all from MSLQ), aimed grade goals, minimum grade goals, performance self-efficacy	Third-year education and child studies (ECS) students & second-year psychology (PSY) students 1. n_{ECS} = 85 n_{psy} = 219 2. & 3. n_{ECS} = 51 n_{psy} = 150	1. T-tests & chi-squared test 2. MANOVA & follow-up ANOVAs 3. Hierarchical multiple regression	b, c & d

Table 1. (Continued)

Ch.	Aim(s)/RQs	Measures	Sample	Analyses	Paths
5	1. Which latent classes of students exist in terms of the development of grade goals and performance self-efficacy throughout the first academic year? 2. How does latent class-membership relate with students' course evaluations and academic performance?	1. Grade goals & performance self-efficacy in eight single courses 2. Students' course evaluation items & academic performance.	First-year social sciences students: Psychology (*n* = 349), international psychology (*n* = 136), education and child studies (*n* = 102)	1. Latent class analysis 2. T-tests	b & c
6	Why is student motivation in higher education sensitive to assessments?				a, b & c

Note. Ch. = Chapter; RQ = Research Question; the numbers in the measures, samples and analyses columns correspond with the number of the aim; the paths column indicates which paths of Figure 1 are studied per chapter.

CHAPTER 2

Assessment policies and academic progress: Differences in performance and selection for progress

This chapter is submitted as:

Kickert, R., Meeuwisse, M., Arends, L.R., Prinzie, P., & Stegers-Jager, K.M (submitted). Assessment policies and academic progress: Differences in performance and selection for progress

Abstract

Despite the benefits swift academic progress holds for many stakeholders, there is scarce literature on how academic progress may be improved by changes to assessment policies. Therefore, we investigated academic progress of first-year students after an alteration of characteristics of the assessment policies in three large course programmes: Business Administration (n = 2048) changed the stakes; Medicine (n = 1630) changed the stakes and performance standard; Psychology (n = 1076) changed the stakes, performance standard and resit standard. Results indicate that students' academic progress was sensitive to the characteristics of the assessment policy in all three course programmes. The changes in progress could be explained by differences in performance (e.g. GPA), as well as by differences in selection for progress by the different policies. Implications are that assessment policies seem effective in shaping student progress, although one size does not fit all.

Keywords: assessment policies, academic progress, academic performance, stakes, performance standards, resit

Introduction

Swift academic progress for many students saves time, money and energy for students, educators, as well as society. Therefore, optimising academic progress is an important goal for educational stakeholders worldwide (Attewell et al., 2011; Vossensteyn et al., 2015). Adapting characteristics of assessment policies may be an efficient way to improve academic progress, given the premises that (i) characteristics of assessment policies are related with student grades (Cole & Osterlind, 2008; Elikai & Schuhmann, 2010; Kickert et al., 2018), and (ii) decisions about academic progress are based on students' grades. Recently, in an attempt to accelerate first-year academic progress, three large faculties of a large Dutch university changed their assessment policies. This change created a rare natural quasi-experiment, which lends an opportunity to investigate how assessment policies affect academic progress.

Assessment Policies

We define an *assessment policy* as the organisational structure of assessments within a course programme. This policy also describes the criteria that are utilised to decide about students' academic progress. In this study, we use the term *academic progress* to denote whether a student has obtained all credits of the first year of the course programme. In the current investigation, we will compare academic progress under assessment policies that differ on three characteristics: (i) the height of the stakes, (ii) the performance standard, and (iii) the resit standard.

Height of the Stakes

The *height of the stakes* refers to the consequences of failing one or more assessments. In Dutch higher education, first-year students need to progress to the second year within a fixed timeframe, in order to avoid academic dismissal (Arnold, 2015). Therefore, in the current investigation, the height of the stakes is determined by the length of this timeframe. For instance, the consequences of failing one or more assessments are higher when first-year students are required to progress within one year instead of two years.

The published studies on the relationship between the stakes and academic progress show mixed results. On the one hand, it has been shown that higher stakes on single tests are associated with higher grades (Cole & Osterlind, 2008; Wolf & Smith, 1995). Consequently, raising the stakes might be an efficacious way to enhance academic progress. Research on academic probation shows that setting a minimum standard for future performance of low-performing students, encourages some students to drop out, while improving grades for those students who decide to stay in the

course programme (Lindo et al., 2010). On the other hand, previous research on Dutch assessment policies showed higher first-year dropout rates (Arnold, 2015; Sneyers & De Witte, 2015), as well as lower grades (De Koning et al., 2014) under academic dismissal policies. Additionally, results on academic progress were mixed, showing either no increase in progress (Eijsvogels et al., 2015; Stegers-Jager et al., 2011), or even a slight decrease in obtained credits (De Koning et al., 2014) after the introduction of an academic dismissal policy. However, in these previous investigations, assessment policies with a two-year timeframe for progress were compared with policies without a timeframe requirement for progress. In the current investigation we compared one-year timeframe policies with two-year timeframe policies. In other words, research hitherto has compared high stakes to low stakes, whereas in the current study we compare high stakes to even higher stakes.

Performance Standard

The *performance standard* refers to the minimum grade standard for the assessment of a course, to obtain the corresponding course credits. Thus, performance standards specify which grades result in academic progress. With compensatory performance standards, decisions on academic progress are based upon the average grade, thus allowing compensation of lower grades with higher grades. In case of conjunctive performance standards, students need to pass each individual assessment with a satisfactory grade (Chester, 2003).

On the one hand, higher performance standards have consistently been associated with higher grades (Elikai & Schuhmann, 2010; Johnson & Beck, 1988; Kickert et al., 2018, 2019), which should result in higher progress. Additionally, simulation studies have shown that more students progress in case of compensatory instead of conjunctive standards (Douglas & Mislevy, 2010; Yocarini et al., 2018). On the other hand, a higher performance standard is harder to pass, which may result in lower progress (Yocarini et al., 2018). Due to these two opposing influences of higher performance standards on academic progress, it is difficult to predict whether progress will be affected by an altered performance standard in real life. To the best of our knowledge, no real-life observational research on the effects of performance standards on progress is available, possibly due to the rarity of an alteration of the performance standard of an entire assessment policy.

Resit Standard

The *resit standard* refers to the number of permitted resits. Firstly, resit standards can be adjusted by only allowing for a portion of the courses to be retaken. Secondly, constraints can be put on the number of times each assessment can be retaken.

Simulation studies on resits suggest that more resits will result in higher academic progress in two ways (Douglas & Mislevy, 2010). Firstly, students may increase their true ability before a next attempt (McManus, 1992). Secondly, resits can unfortunately also offer an unfair opportunity to students who have not yet attained a proper level, but may still pass a test by chance (Yocarini et al., 2018). However, these simulation studies did not capture alterations in student performance due to different resit standards. Empirical evidence on student grades shows that a higher number of allowed resits is related with lower grades on the initial assessment, but not related with final grades (Grabe, 1994). In that case, academic progress should also be unaffected by a different resit standard. To the best of our knowledge, there are no previous empirical investigations of the association between resit standards and academic progress.

Two Ways From Assessment Policies to Academic Progress

In this study we focused on the height of the stakes, the performance standard and the resit standard as the key characteristics of assessment policies. We examined academic progress under assessment policies that differ in terms of these three characteristics. We distinguished between two possible ways in which assessment policies may influence academic progress. Firstly, assessment policies may affect *performance*. Changing the assessment policy may cause students to study differently, and consequently result in differences in academic performance. For example, higher stakes and performance standards have been associated with better self-regulated learning, more participation in scheduled learning activities and higher grades (Kickert et al., 2018). Thus, different assessment policies may cause differences in performance, which in turn could result in differences in academic progress.

Secondly, changing the assessment policy may result in a different *selection for progress* of first-year students who will progress to the second year (Douglas & Mislevy, 2010; Yocarini et al., 2018). As assessment policies specify the relationship between grades and progress, grades that would lead to progress under one assessment policy, may not lead to progress under another policy. Thus, the pool of students that is selected for progress will be different under different assessment policies

In sum, when changes to assessment policies are made, performance and selection for progress are expected to change simultaneously. Due to this simultaneous change of performance and selection, in practice it is difficult to separate the influences that performance and selection for progress have on academic progress under different assessment policies. However, if academic progress increases, it is important to understand whether this happened because students are showing improved performance, or because the selection has become easier. Therefore, in the current

study we attempted to monitor differences in performance and differences in selection for progress under different assessment policies.

Research Questions

In the current investigation, we aimed to answer three research questions (RQs): 1) What is the relationship between differences in assessment policies and differences in academic progress?; 2) What is the relationship between differences in assessment policies and differences in performance?; 3) What is the relationship between differences in assessment policies and differences in selection for progress? For RQ1, we compared academic progress under an old lower-stakes assessment policy versus a new higher-stakes policy in three course programmes. In order to answer RQ2, we first investigated differences in average academic performance, i.e. Grade Point Average (GPA (RQ2a)). In addition, we obtained a second performance indicator: we mimicked whether students would have progressed if they had studied under the performance standards of a different assessment policy (RQ2b). Then, performance is not only operationalised as average grades, but also as whether the performance meets different standards: Well-performing students should progress under different performance standards as well. In order to answer RQ3, we also used students' mimicked academic progress, to see whether the selection for progress differs between assessment policies.

Methods

Curricula and Assessment Policies

Data were gathered at a large urban university in the Netherlands at three course programmes that changed their assessment policies in order to accelerate academic progress: Business Administration, Medicine and Psychology. In all three course programmes, the three-year bachelor's programme consists of 60 credits per year. First-year students who drop out before February 1st are allowed to re-enter the same programme at the start of the next academic year. Moreover, these early drop-outs need not reimburse their student loans.

The three course programmes changed their assessment policies in different academic years: Psychology switched in 2011, Business Administration in 2012, and Medicine did so in 2014. In Table 1, a schematic overview of the characteristics of the *lower-stakes* (old) and *higher-stakes* (new) assessment policies per course programme is provided. In all three course programmes, the stakes were adapted similarly; under the *lower-stakes* assessment policies, first-year students needed 40 first-year credits within one

year to evade academic dismissal[1], and all 60 first-year credits within two years; in the *higher-stakes* assessment policies, all 60 credits need to be obtained within one year in order to evade academic dismissal. For Business Administration, the main adaptation to the assessment policy was the change in stakes. Medicine changed the stakes as well as the performance standard. Psychology adapted the stakes, the performance standard and the resit standard. We should note, that in the lower-stakes Psychology policy the performance standard and resit standard were different for the Skills assessments and Knowledge Assessments[2]. Detailed descriptions of the three course programmes, as well as the respective changes to the three assessment policies, can be found in Appendix 1.

Participants

There were two inclusion criteria for the current study. Firstly, to assure we would only use students who were affected by the assessment policy, students needed to have obtained at least one grade. Secondly, we excluded students who had previously been enrolled in the same course programme, as these students may have obtained grades under two different assessment policies. For each course programme, we compared the last two cohorts of first-year students under the lower-stakes assessment policy (i.e. *lower-stakes policy students*) with the first two first-year cohorts under the higher-stakes policy (i.e. *higher-stakes policy students*), resulting in a total of n = 4754 students. However, for Business Administration we only used the final (2011) cohort under the lower-stakes policy, as the introduction of a goal-setting intervention one year before the change in stakes (see (Schippers et al., 2015) could confound our results. Thus, for Business Administration we compared the cohort of 2011 from the lower-stakes assessment policy (n = 656, 72.1% male, M_{AGE} = 18.8 years, SD_{AGE} = 1.2 years), to cohorts 2012 and 2013 from the higher-stakes assessment policy (n = 1392, 68.5% male, M_{AGE} = 18.7 years, SD_{AGE} = 1.2 years). For Medicine, we compared the cohorts of 2012 and 2013 from the lower-stakes assessment policy (n = 809, 37.9% male, M_{AGE} = 19.5 years, SD_{AGE} = 2.1 years) with cohorts 2014 and 2015 from the higher-stakes policy (n = 821, 33.6% male, M_{AGE} = 19.2 years, SD_{AGE} = 2.0 years). For Psychology we compared the cohorts of 2009 and 2010 for the lower-stakes policy (n = 558, 25.3% male, M_{AGE} = 19.9 years, SD_{AGE} = 3.3 years), to those of 2011 and 2012 for the higher-stakes assessment policy (n = 518, 26.3% male, M_{AGE} = 19.7 years, SD_{AGE} = 2.4 years).

1 In Medicine, only students with less than 40 credits who failed to attend compulsory support meetings were dismissed.

2 Within the Psychology Curriculum, a distinction exists between Knowledge courses and Skills trainings, please see Appendix 1 for more information.

Table 1. *The lower-stakes and higher-stakes assessment policies of the three course programmes currently under study.*

		Business administration		**Medicine**		**Psychology**			
		Lower-stakes (cohort 2011)	*Higher-stakes (cohorts 2012 & 2013)*	*Lower-stakes (cohorts 2012 & 2013)*	*Higher-stakes (cohorts 2014 & 2015)*	*Lower-stakes (cohorts 2009 & 2010)*		*Higher-stakes (cohorts 2011 & 2012)*	
Height of the stakes	*1 year credit requirement*	40	60	40	60	40		60	
	2 year credit requirement	60	-	60	-	60		-	
						Knowledge	*Skills*	*Knowledge*	*Skills*
Performance standard	*N compensable grades (n courses)*	1 (12)	1 (12)	0 (9)	2 (9)	2 (8)	0 (9)	8 (8)	9 (9)
	Lowest compensable grade allowed	4.5	4.5	-	5.0	1.0	-	4.0	4.0
	Minimum GPA	5.5	5.5	-	6.0	6.5 (semi-formative)	-	6.0	6.0
	Lowest conjunctive grade allowed	5.5	5.5	5.5	5.5	5.5	5.5	-	-
Resit standard	*Maximum allowed number of courses*	4	4	9	9	0	9	2	2
	Final grade	Latest	Latest	Highest	Highest	-	Highest	Highest	Highest

Table 1. *(Continued)*

	Business administration		Medicine		Psychology		
Other changes	-	Slight changes to form of assessments	Minor changes in the distribution of credits over courses	Minor changes in the distribution of credits over courses	-	- Progress tests no longer used	- 1 more credit for the 9th skills training; 40 instead of 41 credits for all knowledge assessments

Note. Grades for separate assessments are given on a scale from 1 (lowest score) to 10 (perfect score). In case of compensatory assessment policies where not all grades can be compensated, the 'lowest conjunctive grade allowed' entails the threshold below which grades need to be compensated. Semi-formative indicates that lower-stakes policy Psychology students could progress on the basis of the knowledge assessments, but were not required to do so; progress tests were the principal way to progress. See Appendix 1 for a detailed description.

Measures

For Business Administration and Psychology, all data were obtained from the Erasmus Education Research Database. For Medicine, the data were not yet available in the database, and thus were obtained from the university student administration system.

Academic Progress

Actual progress. We operationalised actual academic progress as students obtaining all 60 first-year credits of the course programme within the set timeframe. In the lower-stakes assessment policy, students could take a maximum of two years to progress; in the higher-stakes policy, students only get one year. Therefore, from this point on we will differentiate between one-year progress, and final progress. In the higher-stakes assessment policies, one-year progress is identical to final progress.

Mimicked progress. In addition to the actual academic progress, we mimicked whether each student would have progressed under the performance standard of the other assessment policy. More specifically, for lower-stakes policy students we mimicked their academic progress under the performance standards of the higher-stakes policy, and vice versa for higher-stakes policy students. This mimic could only be performed for Medicine and Psychology, since the performance standard did not change for Business Administration students. To determine this mimicked progress, we used students' final grades. These grades were used in reality to determine students' final progress; after two years in the lower-stakes policy, and after one year in the higher-stakes policy. Only students who faced personal circumstances were sometimes exempted from academic dismissal and could thus have obtained grades after these deadlines. Nevertheless, we only used grades after two years in the lower-stakes policy, and after one year in the higher-stakes policy.

Grade Point Average (GPA)

We calculated GPA as the weighted average of the final grades for all students who had at least one first-year grade. Grades for separate assessments are always given on a scale from 1 (lowest score) to 10 (perfect score). All grades were taken into account, regardless of the fact whether the grades were sufficient or not. In Medicine and Psychology, minor changes were made to the distribution of credits over the separate courses (e.g. a course gaining 1 credit at the expense of another course); therefore, we calculated GPA per cohort, weighing the courses appropriately per cohort.

For Business Administration students, the GPA is the average grade on all 12 first-year courses. For Medicine, the GPA is the average grade on nine knowledge assessments;

the skills training assessments are mostly pass/fail-graded and therefore not included in the calculation of the GPA. Psychology students get a separate Knowledge GPA for eight knowledge assessments and a Skills GPA for nine practical assessments.

Statistical analyses

To investigate the differences in *academic progress* under the lower-stakes and higher-stakes assessment policies for all three course programmes (RQ1), we performed chi-squared tests on the number of students who showed academic progress. As lower-stakes policy students could take two years to progress, for each course programme we performed chi-squared tests on both one-year academic progress and final academic progress under the lower-stakes and higher-stakes assessment policy. We included odds ratios as measures of effect size (1.22/1.86/3.0 = small/medium/large; or inverse equivalents 0.82/0.54/0.33 = small/medium/large; Olivier & Bell, 2013).

In order to clarify how differences in assessment policies relate to differences in *performance* (RQ2), we performed two analyses. Firstly, we compared the GPA between the lower-stakes and the higher-stakes policies (RQ2a). We performed two t-tests on GPA: a t-test comparing all lower-stakes policy vs. all higher-stakes policy students, and a t-test comparing only the students who progressed under the lower-stakes vs. the higher-stakes policy. We calculated Cohen's *d* as a measure of effect size (.20/.50/.80 = small/medium/large effect size; Cohen, 1992).

As a second performance measure, we mimicked whether students would have progressed under the performance standards of the lower-stakes as well as the higher-stakes policy (RQ2b). Progress could only be mimicked for Medicine and Psychology, as Business Administration did not alter the performance standard. Therefore, we performed two chi-squared tests for the differences in mimicked progress for lower-stakes policy versus higher-stakes policy Medicine and Psychology students, under the performance standards of: i) the lower-stakes assessment policy, and ii) the higher-stakes assessment policy. If a group of students shows higher progress under their own policy, as well as under the alternative policy, this indicates that these students perform better than the other group of students. Additionally, if students show higher progress under their actual performance standards, compared to the alternative performance standards, this indicates that these students' performance is sensitive to the performance standard. We calculated odds ratios as measures of effect size (Field, 2013).

Finally, we tested whether the *selection* made by the performance standards of the lower-stakes and higher-stakes assessment policies of Medicine and Psychology

differed (RQ3), by performing McNemar tests on the association between students' mimicked progress under the lower-stakes and higher-stakes policies. We performed three separate tests: for all students together, for lower-stakes policy students and for higher-stakes policy students. If the selection is different under different performance standards, students would show progress under one policy, but not under the other.

Results

Academic Progress (RQ1)

We first investigated differences in actual academic progress under the lower-stakes vs. the higher-stakes policy for each course programme (RQ1). For *Business Administration*, one-year progress in the higher-stakes assessment policy was significantly higher than in the lower-stakes policy, $\chi^2(1) = 79.01$, $p < .001$, $OR_{\text{progress lower-stakes /higher-stakes}} = 0.41$. Final progress in the higher-stakes policy was significantly lower than final progress in the lower-stakes policy, $\chi^2(1) = 24.59$, $p < .001$, $OR_{\text{progress lower-stakes /higher-stakes}} = 1.62$. See Table 2 for the descriptives of the study variables for Business Administration.

For *Medicine*, students in the higher-stakes assessment policy showed significantly higher one-year progress than students in the lower-stakes policy, $\chi^2(1) = 70.00$, $p < .001$, $OR_{\text{progress lower-stakes /higher-stakes}} = 0.42$. However, final progress was significantly lower in the higher-stakes policy than in the lower-stakes policy, $\chi^2(1) = 49.73$, $p < .001$, $OR_{\text{progress lower-stakes /higher-stakes}} = 2.40$. See Table 3 for the descriptives for Medicine.

Psychology students' one-year progress in the higher-stakes assessment policy was significantly higher than one-year progress in the lower-stakes policy, $\chi^2(1) = 61.30$, $p < .001$, $OR_{\text{progress lower-stakes /higher-stakes}} = 0.36$. Final progress was also significantly higher in the higher-stakes policy than in the lower-stakes policy, $\chi^2(1) = 4.59$, $p = .032$, $OR_{\text{progress lower-stakes /higher-stakes}} = 0.75$. See Table 4 for the descriptives for Psychology.

Table 2. *Descriptives for Business Administration: academic progress (RQ1) and performance (RQ2a) of students under the lower-stakes and higher-stakes assessment policy*

	Real Progress (RQ1)		**GPA (RQ2a)**	
	One-year	**Final**	M_{Total} **(*SD*)**	$M_{Progress}$ **(*SD*)**
Lower-stakes policy students (*N* = 656)	31.4%	64.0%	6.52 (1.02) *N*=656	7.07 (0.51) *N*=420
Higher-stakes policy students (*N* = 1392)	52.4%	52.4%	6.41 (1.19) *N*=1392	7.15 (0.56) *N*=729

Table 3. *Descriptives for Medicine: academic progress (RQ1), performance (RQ2a&b) and selection for progress (RQ3) of students under the lower-stakes and higher-stakes assessment policy.*

	Real Progress (RQ1)		**GPA (RQ2a)**		**Mimicked Progress (RQ2b)**		**Selection for progress (*N*) (RQ3)**			
	One-year	**Final**	M_{Total} **(*SD*)**	$M_{Progress}$ **(*SD*)**	**LSP P.S.**	**HSP P.S.**				
Lower-stakes policy students (*N* = 809)	50.9%	85.5%	6.38 (.85) *N*=805	6.62 (.56) *N*=692	85.7%	80.6%		Progress	LSP P.S.	
							Progress		No	Yes
							HSP	No	110	47
							P.S.	Yes	6	646
Higher-stakes policy students (*N* = 821)	71.1%	71.1%	6.31 (1.08) *N*=818	6.82 (.65) *N*=584	48.4%	71.0%		Progress	LSP P.S.	
							Progress		No	Yes
							HSP	No	237	1
							P.S.	Yes	187	396

Note. LSP = Lower-stakes policy; HSP = Higher-stakes policy; P.S. = Performance standard

Differences in Performance (RQ2)

Differences in GPA (RQ2a)

Subsequently, we investigated differences in GPA under the two assessment policies for each course programme. For *Business Administration*, lower-stakes policy students had a significantly higher GPA than higher-stakes policy students, $t(1480.35) = 2.17$, $p = .030$, $d = .10$. After selecting only the (final) progressing students, lower-stakes

policy students showed a significantly lower GPA than higher-stakes policy students, $t(940.12) = -2.45$, $p = .014$, $d = -.15$ (Table 2).

For *Medicine*, we did not find a statistically significant difference between the GPA of all lower-stakes policy students and all higher-stakes policy students, $t(1551.38) = 1.46$, $p = .143$, $d = .07$. When comparing the GPA of progressing students, lower-stakes policy students showed a significantly lower GPA than higher-stakes policy students, $t(1159.16) = -5.92$, $p < .001$, $d = -.34$ (Table 3).

For *Psychology*, when comparing all students, the Knowledge GPA was significantly lower under the lower-stakes policy than under the higher-stakes policy, $t(1067.30) = -6.20$, $p < .001$, $d = -0.38$. However, the Skills GPA was significantly higher for lower-stakes policy students than for higher-stakes policy students, $t(868.30) = 6.60$, $p < .001$, $d = 0.41$. After selecting the progressing students, lower-stakes policy students still showed a significantly lower Knowledge GPA than higher-stakes policy students, $t(701.45) = -7.21$, $p < .001$, $d = -0.52$. Again, the Skills GPA was significantly higher for lower-stakes policy students than for higher-stakes policy students, $t(746.69) = 5.61$, $p < .001$, $d = 0.40$ (Table 4).

Differences in Mimicked Progress (RQ2b)

Next, for Medicine and Psychology we compared lower-stakes versus higher-stakes policy students' mimicked progress under the performance standards of the lower-stakes policy, as well as under the higher-stakes policy. For *Medicine*, under the performance standards of the lower-stakes assessment policy, lower-stakes policy students showed significantly higher progress than higher-stakes policy students, $\chi^2(1) = 255.98$, $p < .001$, $OR_{\text{progress lower-stakes /higher-stakes}} = 6.38$. Under the performance standards of the higher-stakes assessment policy, lower-stakes policy students also showed significantly higher progress than higher-stakes policy students , $\chi^2(1) = 20.38$, $p < .001$, $OR_{\text{progress lower-stakes /higher-stakes}} = 1.70$. Thus, compared to higher-stakes policy students, lower-stakes policy Medical students showed higher progress under both the lower-stakes and the higher-stakes performance standards (Table 3).

Table 4. *Descriptives for Psychology: academic progress (RQ1), performance (RQ2a&b) and selection for progress (RQ3) of students under the lower-stakes and higher-stakes assessment policy.*

	Real Progress (RQ1)		**GPA (RQ2a)**				**Mimicked Progress (RQ2b)**		**Selection for progress (*N*) (RQ3)**			
	One-year	**Final**	**$M_{knowledge\text{-}total}$ (*SD*)**	**$M_{skills\text{-}total}$ (*SD*)**	**$M_{knowledge\text{-}progress}$ (*SD*)**	**$M_{skills\text{-}progress}$ (*SD*)**	**LSP P.S.**	**HSP P.S.**				
Lower-stakes policy students (*N* = 558)	51.6%	68.8%	5.92 (1.25) *N*=554	7.20 (0.55) *N*=556	6.42 (0.91) *N*=384	7.37 (0.38) *N*=384	34.2%	45.9%		Progress LSP P.S.		
									Progress		No	Yes
									HSP	No	300	2
									P.S.	Yes	67	189
Higher-stakes policy students (*N* = 518)	74.7%	74.7%	6.37 (1.12) *N*=517	6.91 (0.86) *N*=517	6.83 (0.67) *N*=387	7.21 (0.45) *N*=387	36.7%	74.9%		Progress LSP P.S.		
									Progress		No	Yes
									HSP	No	130	0
									P.S.	Yes	198	190

Note. LSP = Lower-stakes policy; HSP = Higher-stakes policy; P.S. = Performance standard

For *Psychology,* under the performance standards of the lower-stakes assessment policy, lower-stakes policy students' progress did not differ significantly from higher-stakes policy students' progress, $\chi^2(1) = 0.71$, $p = .401$, $OR_{progress\ lower\text{-}stakes\ /higher\text{-}stakes} = 0.90$. Under the performance standards of the higher-stakes assessment policy, lower-stakes policy students showed significantly lower progress than higher-stakes policy students, $\chi^2(1) = 94.18$, $p < .001$, $OR_{progress\ lower\text{-}stakes\ /higher\text{-}stakes} = 0.28$. Thus, compared to lower-stakes policy students, higher-stakes policy Psychology students only showed higher progress under the higher-stakes performance standards (Table 4).

Differences in Selection for Progress (RQ3)

Finally, we investigated differences in selection for progress between the lower-stakes and higher-stakes policies (RQ3), i.e. how many students would progress under one policy but not the other. For *Medicine*, the lower-stakes and the higher-stakes policy differed significantly in which students would have been selected for progress, $\chi^2(1) = 86.04$, $p < .001$. These differences between both policies also hold true when comparing the selection for progress separately for lower-stakes policy students, $\chi^2(1) = 30.19$, $p < .001$, and for higher-stakes policy students, $\chi^2(1) = 182.05$, $p < .001$. For Medicine, 15% of students would progress under one policy but not the other. Lower-stakes policy students would show higher progress under the performance standard of the lower-stakes policy, whereas the opposite pattern emerged for higher-stakes policy students (Table 3).

For *Psychology*, the lower-stakes and the higher-stakes policy also differed significantly in which students would have been selected for progress, $\chi^2(1) = 257.09$, $p < .001$. These differences between both policies also hold true when comparing the selection for progress separately for lower-stakes policy students, $\chi^2(1) = 59.36$, $p < .001$, and for higher-stakes policy students, $\chi^2(1) = 196.01$, $p < .001$. For Psychology, 25% of students would only progress under one of the policies. Both lower-stakes policy and higher-stakes policy students would progress more under the performance standard of the higher-stakes policy (Table 4).

Conclusion and Discussion

The current investigation aimed to clarify possible differences in academic progress (RQ1), academic performance (RQ2) and selection for progress (RQ3) after alterations to characteristics of assessment policies in three course programmes: only the stakes were adapted in the Business Administration policy, in Medicine both the stakes and the performance standard were changed, and in Psychology the stakes, performance standard and resit standard were altered. Overall, we can conclude that students' progress is associated with characteristics of the assessment policy, and this association can be explained by differences in performance, as well as by differences in selection for progress by the different policies. Below we will discuss our findings in more detail, for progress, performance and selection for progress separately.

Differences in Academic Progress

In terms of academic progress (RQ1), in all three faculties we observed significantly higher one-year progress in the higher-stakes assessment policies compared with the lower-stakes policies. Thus, progress was faster in case of higher stakes, i.e. when students were required to obtain all first-year credits within one year, instead of two years. This means that many students seem to adapt their pace of progress to the demands of the assessment policy. However, we found mixed results for final progress, which was measured after two years in the lower-stakes policies and after one year in the higher-stakes policies: final progress in the higher-stakes policy was lower in Business Administration and Medicine, yet higher in Psychology.

In other words, for Business Administration and Medicine, academic progress in the higher-stakes assessment policies was faster than in the lower-stakes policy, as more students progressed after one year. However, final progress was lower. Thus, a large share of the higher-stakes policy students seems to have adapted the pace of their academic progress to the requirement of obtaining all 60 credits within one year, but not all students were able to do so within this shorter timeframe of the higher-stakes policy. The fact that final progress was lower in the higher-stakes policy for Business Administration and Medicine, may indicate a ceiling effect; some students may not be able to progress within one year in these course programmes (Stegers-Jager & Themmen, 2015). This ceiling effect is particularly relevant in Medicine, where final progress was already high in the old policy. Conversely, Psychology students did show higher final progress in the higher-stakes policy as well, which suggests the absence of a ceiling effect here. Thus, for Psychology students, progress was faster and higher under the higher-stakes assessment policy.

The lower final progress in Business Administration and Medicine is somewhat consistent with previous investigations on academic dismissal policies, which either found no difference (Eijsvogels et al., 2015; Stegers-Jager et al., 2011), or a decrease in obtained first-year credits (De Koning et al., 2014). However, the higher final progress in Psychology, and the higher one-year progress in all three course programmes, is not in line with these previous studies. This discrepancy between our and previous findings can be explained by the fact that previous investigations on academic dismissal made a comparison between low stakes and high stakes, i.e. an unlimited timeframe versus a two-year timeframe, respectively. Contrarily, we compared high to even higher stakes, i.e. a two-year versus a one-year timeframe, respectively.

Differences in Performance

Overall, based on our results we can conclude that assessment policies matter for performance (RQ2), and may therefore offer part of the explanation of the differences in progress under the different policies. Again, we should note that lower-stakes policy students could take two years to attain their final grades, compared to only one year for higher-stakes policy students. For Business Administration, where only the stakes were changed, lower-stakes policy students outperform higher-stakes policy students when comparing the GPA (RQ2a) of all students. However, results are inversed when only progressors under both policies are compared. An explanation can be found in the lower final progress rate under the higher-stakes policy, which indicates that the higher-stakes policy may be more selective. Progressors' higher grades under the higher-stakes policy might be a consequence of this selectivity.

These inversed results for all students versus progressors underline the importance of choosing the appropriate population of interest in evaluating the consequences of policy changes. In this case, as the progressing students remain potential graduates, we feel that this is the subpopulation of students for whom it is particularly relevant to improve performance. In essence, a student who progresses with better performance, should be a better graduate as well. Only comparing all students would have obscured the differences between progressors under both policies. Thus, educators will have to make a context-specific decision about which student groups are most relevant to compare.

For Medicine, where the stakes and the performance standard were adapted, only the progressing students obtained higher GPAs in the higher-stakes policy (RQ2a). Again, the explanation may be the lower final progress rate under the higher-stakes policy, which indicates that the higher-stakes policy may be more selective, resulting in higher grades for progressors. Mimicking Medical students' progress under the alternative

assessment policy (RQ2b), indicated that lower-stakes policy students would have progressed more, under both the lower-stakes and higher-stakes policy performance standards. This higher mimicked progress points towards superior performance for lower-stakes policy students. Concluding, although progressing Medical students' GPA is better under the higher-stakes assessment policy, the mimicked performance indicator in Medicine implies better performance under the lower-stakes policy. This discrepancy underlines the importance of the type of performance indicator chosen to evaluate the consequences of policy changes. For instance, GPA is less relevant under policies with conjunctive performance standards; a student with a good GPA may have failed one or more individual courses, and thus this student's performance is insufficient to progress.

In the Psychology assessment policy, the stakes, performance standard and resit standard were adapted. Here, a contrasting picture emerged for the Knowledge and Skills GPA (RQ2a); in the higher-stakes policy, the Knowledge GPA was higher, but the Skills GPA was lower. Different from Business Administration and Medicine, this pattern was similar when comparing only progressing students. Results from the mimicked progress (RQ2b), indicate better performance under the higher-stakes policy, as higher-stakes policy Psychology students outperformed lower-stakes policy students under both the lower-stakes and higher-stakes performance standards. Thus, again we observe that the choice of performance measure matters for the conclusions: lower-stakes policy students outperform higher-stakes policy students in terms of Skills GPA, but the reversed is true based on the Knowledge GPA and mimicked progress.

Although we cannot currently draw any causal conclusions, we expect the discrepancy between the Knowledge and Skills GPAs of Psychology students to result from a combination of factors. Firstly, although the higher-stakes performance standards were identical for the Knowledge and Skills assessments, the lower-stakes performance standards were not. For instance, the Knowledge assessments were semi-formative in the lower-stakes policy and summative in the higher-stakes policy, whereas the Skills assessments were summative in both policies. Additionally, the average lower-stakes policy Knowledge GPA was below the higher-stakes policy performance standard, while the average lower-stakes policy Skills GPA was more than two standard deviations above the higher-stakes performance standard. Consequently, a rise in Knowledge GPA was necessary to meet the higher-stakes standards, and more salient because the stakes of the knowledge assessments were raised. Thirdly, the type of assessed learning (e.g. Knowledge or Skills) may matter for the consequences of altered assessment policies. Thus, from the performance of Psychology students we can conclude that assessment policies may shape student performance, but that

other factors such as the type of assessed learning may affect the consequences of the choices made.

Generally speaking, in terms of performance the results are in line with previous literature. For progressing Business Administration and Medical students, as well as for the Knowledge assessments in Psychology, we replicated results on higher stakes' association with better grades (Cole & Osterlind, 2008; Wolf & Smith, 1995). Additionally, our observation of higher GPAs for progressing students in the higher-stakes policy is in line with literature on academic probation, which shows that when the stakes are higher, drop-out is higher but performance of remaining students is also better (Lindo et al., 2010). As we found performance differences in all three course programmes, it seems that assessment policies can effectively be used to shape student performance. However, the divergence between course programmes, between types of assessment, and between performance measures underlines the importance of taking context into account when evaluating policy changes.

Differences in Selection for Progress

Our investigation of differences in selection by the different assessment policies (RQ3) showed that the assessment policies in Medicine and Psychology made different selections for progress; significant numbers of students would progress under one policy but not the other. Thus, there were significant numbers of students for whom it mattered which performance standard was used. Therefore, in addition to differences in performance, differences in selection for progress seem to be a factor in the observed changes in academic progress under the different assessment policies.

The lower-stakes policies seemed to be stricter, as there were more students in both Medicine and Psychology who would progress under the higher-stakes policy but not the lower-stakes policy, than vice versa. For Medicine, relatively more lower-stakes policy students would only progress under the lower-stakes policy than only under the higher-stakes policy; and relatively more higher-stakes policy students would only progress under the higher-stakes policy. In other words, it seemed that Medical students adapted their performance to the standards of the assessment policy. For Psychology we found a different pattern: both lower-stakes and higher-stakes policy students would progress more under the higher-stakes policy. It makes intuitive sense that lower-stakes policy students did not adapt their performance to the lower-stakes performance standards, as these standards were semi-formative, and therefore not salient for students. Alternatively, it may be the case that the lower-stakes policy in Psychology was simply more difficult.

Our results add to the existing body of knowledge on differences in selection by assessment policies, because previous investigations were simulation studies in which the necessary assumption was made that students behave similar under different standards (Yocarini et al., 2018). The current study shows that this is not a realistic assumption, as student performance differed significantly under different assessment policies. Consequently, in addition to evaluating the decision accuracy of the assessment policy, the motivating aspects of the policy need consideration as well.

Limitations and Strengths

This research has several limitations that need to be addressed. Firstly, through observational research, it is impossible to draw any causal conclusions; other factors may affect the observed associations. For instance, other minor changes to courses may have been made in the interval that we investigated. It is particularly important that the assessments in the three course programmes have remained comparable. We believe this to be the case, due to the existence of an examination board in all three course programmes. These examination boards are responsible for the quality of the assessments, as well as for determining the pass/fail-score per assessment. Despite this limitation on causal conclusions, observational research adds unique value, as the importance that academic progress holds for most students cannot be prompted in an experimental setting.

Another limitation is that changes to the stakes, performance standards and resit standards of the assessment policies, were made simultaneously. Therefore, it is impossible to unravel the isolated effects of these characteristics of the assessment policy. For this reason, we chose one course programme that only altered the stakes (i.e. Business Administration), one course program in which both stakes and performance standard were adapted (i.e. Medicine), and one course program in which the stakes, performance standard and resit standard were adjusted (i.e. Psychology). Comparing the conclusions for the three different programmes can only give tentative insights into the isolated effects of changing the stakes: students seem highly sensitive to changes in the stakes in all three programmes. Besides this tentative conclusion on the stakes, we intentionally refrained from making comparisons between the course programmes, as the three programmes are bound to have other differences besides those of the assessment policies between them as well. For instance, the student populations differ substantially between the programmes in terms of gender. However, it seems that for each course programme, assessment policies matter for progress, performance and selection for progress.

Implications and Future Directions

The results raise several issues about the relation of students' progress and performance with characteristics of assessment policies. Firstly, given the significant and substantial differences in progress and performance under different policies, it seems worthwhile to compare progress and performance under a greater variation of assessment policy characteristics. Assuming that performance on assessments is a reflection of learning, adapting the assessment policy has the potential to be a highly effective source of improved learning. It would be particularly interesting to establish the consequences of the alteration of isolated characteristics of assessment policies, instead of the current combinations of changes. For instance, what would happen to performance when only the performance standard is adapted?

Secondly, we should note that based on the current data we cannot draw any conclusions on what amount of progress is the 'right' amount. In other words, we cannot tell whether higher progress under a certain assessment policy is desirable. Perhaps lower progress rates imply a better selection for progress; only students' future performance within and outside the course programme and will tell.

Thirdly, possible negative effects of raising the standards need to be considered as well. Firstly, negative consequences may include a lowered motivation for lifelong learning (Harlen & Crick, 2003). Perhaps a high stakes assessment policy does not adequately prepare students for a life in which setting personally motivating goals is an important skill. Secondly, since assessments are often unable to cover the full range of learning activities (Biggs, 1996; UNESCO, 2016), an increased focus on assessments may lower the time and energy devoted to the unassessed learning activities. Thirdly, students' wellbeing needs to be monitored. Higher standards may raise student stress-levels, which are associated with health problems (Glaser & Kiecolt-Glaser, 2005) and lower academic performance (Akgun & Ciarrochi, 2003). Fourthly, vulnerable groups of students may require special scrutiny, as higher standards may be inequitable for these students.

A final implication is that a careful consideration of the mechanisms by which assessment policies affect performance and selection is necessary. For instance, in terms of motivation, specific, more difficult goals can increase motivation and/or performance (Locke & Latham, 2002), but goals that are too high may lead to failure, and therefore damage self-efficacy (Bandura, 1982). An important question then, is where the tipping point in the relation between goal difficulty and motivation is located. An alternative to this variable-centred approach, is a person-centred approach (Laursen & Hoff, 2006). A question then could be which types of students exist in

terms of sensitivity to the characteristics of the assessment policy. Perhaps some students are strongly focused on meeting the minimum standards of the assessment policy, while other students set their own standards. It would be interesting to see how many students merely want to meet the minimum standards, and how many strive for more.

Conclusion

In conclusion, this study provides empirical evidence that assessment policies are related with academic progress, and this relationship may be explained by differences in performance, as well as differences in selection for progress. Given the apparent tendency of students to perform to the standards of assessment, both in terms of progress and grades, assessment policies seem to be an effective way to shape student progress and performance. Therefore, in addition to evaluating the psychometric properties of an assessment policy, the motivational consequences need careful consideration. The observed differences between course programmes, between different types of assessment within Psychology, as well as between different types of performance indicators within Medicine, underline the importance of a contextualised and nuanced understanding of the relationship between assessment policies, progress and performance; one size does not fit all.

Appendix 1 – Descriptions of Course Programmes and Assessment Policies

Business Administration

The first-year curriculum of Business Administration consists of 12 courses, which have remained stable in terms of content and weighting. Eleven courses are knowledge-focused and one focuses on skills training. The change in *stakes* entailed that under the *lower-stakes* assessment policies, first-year students needed 40 first-year credits within one year to evade academic dismissal, and all 60 first-year credits within two years; in the *higher-stakes* assessment policies, all 60 credits need to be obtained within one year. There were no changes in terms of the *performance standard*: both in the old and higher-stakes policy all individual grades needed to be 5.5 or higher (on a 10-point scale), except for one compensable grade that needed to be at least 4.5. This grade could be compensated by two 7's, or one grade of 8 or higher. Concerning the *resit standard*, one year before the change in stakes was introduced, the maximum number of resits for courses was lowered to four, instead of all twelve courses. Additionally, the introduction of goal-setting intervention in the year before the change in stakes (see Schippers et al., 2015) could confound our investigation. Therefore, in this study we only included the last cohort under the lower-stakes policy, instead of the last two cohorts. The final grade for a course was the latest obtained grade, instead of the highest.

We should note that several activation-enhancing measures coincided with the change in the stakes: more interim assessments, assessments about smaller parts of the courses, and more digital assessments. Teachers were advised and supported to take these measures, but not closely monitored. We therefore do not know the extent to which these measures were actually integrated into the curriculum. In sum, for Business Administration the major change to the assessment policy was the change in the height of the stakes.

Medicine

The Medical school is an integrated and theme-oriented curriculum that has remained stable over the years under investigation. The first year consists of three thematic blocks that contain a total of nine knowledge assessments. The form and content of these assessments have also remained stable over the years. Additionally, each year there are between five and seven skills trainings that are mostly pass/fail-graded. There have been some changes to the structural organisation of these skills trainings, resulting in different numbers of skills trainings per cohort. For consistency,

we changed grades for cohorts that had certain skills trainings graded on a 10-point scale, into dichotomous pass/fail grades.

The change in *stakes* was identical to the change in Business Administration: lower-stakes policy students could take two years to obtain all first-year credits, higher-stakes policy students needed to obtain all credits within one year. The *performance standards* only changed for the nine knowledge assessments: The lower-stakes assessment policy was a conjunctive policy, in which students needed to pass each individual knowledge assessment with a grade of at least 5.5. In the higher-stakes policy, a minimum Grade Point Average (GPA) of 6.0 is required, and two grades between 5.00 and 5.49 can be compensated, on the condition that the two compensated grades have not been obtained in the same thematic block. The remaining grades need to be 5.50 or higher. In both the lower-stakes and higher-stakes assessment policy, students need to pass each skills training, hence no changes in the policy concerning these skills trainings were made. Both in the lower-stakes and higher-stakes policy, resits can be taken for all nine knowledge assessments. However, since students in the lower-stakes policy could take two years to complete the first year, there was a maximum of three resit opportunities per assessment; in the higher-stakes policy, there is only one resit opportunity per assessment. Concluding, compared to the lower-stakes assessment policy, the higher-stakes policy for Medicine has higher stakes and a higher but slightly compensatory performance standard, and no altered resit standard.

Psychology

Since it was founded in 2001, Psychology has been a problem-based learning curriculum (Schmidt, 1994). The first year of the bachelor's programme consists of eight course exams with corresponding knowledge assessments and nine skills trainings with corresponding skills assessments. Until the introduction of the higher-stakes assessment policy in 2011, 41 out of 60 first-year credits could be earned through three knowledge progress tests. These tests covered the content of the first two years of the bachelor's programme and thus assessed long-term knowledge (for a full description, see De Koning et al. 2014).

The *stakes* were adapted similarly as in Business Administration and Medicine; higher-stakes policy students needed to progress within one year, compared to two years for lower-stakes policy students. In terms of *performance standards,* before the change of assessment policy the eight knowledge assessments were semi-formative: students could progress on the basis of the three knowledge progress tests, but if the eight knowledge assessments were passed with a GPA of at least 6.5, this would result

in academic progress as well. These eight assessments were fully compensatory; thus, any grade could be compensated. However, only two grades were allowed to be below 5.5. In the higher-stakes assessment policy, the knowledge progress tests are no longer used; the eight knowledge tests became summative and need to be passed with a first-year GPA of at least 6.0. Grades of 4.0 and up can be compensated, grades below 4.0 are considered invalid. In terms of performance standards for the skills trainings, lower-stakes assessment policy students were required to pass all nine skills trainings with a grade of 5.5. In the higher-stakes policy, a skills training-GPA of 6.0 is required; grades of 4.0 and up can be compensated and a maximum of two skills assessments can be retaken. Concerning the *resit standard*, as the knowledge assessments in the lower-stakes assessment policy were semi-formative, there were no resits; there was only a resit for the knowledge progress tests. In the higher-stakes policy, a maximum of two knowledge assessments can be taken as a resit. For the skills training assessments, all assessments could be retaken in the lower-stakes policy, compared to two assessments in the higher-stakes policy. Concluding, compared to the lower-stakes assessment policy, the new Psychology assessment policy has higher stakes. For the Knowledge assessments, the higher-stakes policy has a lower performance standard with less lenient compensation, and a more lenient resit standard. For the Skills assessments under the higher-stakes policy, the performance standard was raised overall (i.e. the average) yet lowered per assessment; the resit standard was stricter.

CHAPTER 3

The role of the assessment policy in the relation between learning and performance

This chapter has been published as:

Kickert, R., Stegers-Jager, K. M., Meeuwisse, M., Prinzie, P., & Arends, L. R. (2018). The role of the assessment policy in the relation between learning and performance. Medical Education, 52(3), 324–335. https://doi.org/10.1111/medu.13487

Abstract

Background: Optimising student learning and academic performance is a continuous challenge for medical schools. The assessment policy may influence both learning and performance. Previously, the joint contribution of self-regulated learning (SRL) and participation in scheduled learning activities towards academic performance has been reported. However, little is known about the relationships between SRL, participation and academic performance under different assessment policies.

Objectives: The goal of this study was to investigate differences in average scores of SRL, participation and academic performance of students under two assessment policies: i) a conjunctive lower stakes, lower performance standard (old) assessment policy and ii) a compensatory higher stakes, higher performance standard (new) assessment policy. In addition, this research investigated whether the relationships between academic performance, SRL and participation are similar across both assessment policies.

Methods: Year-1 medical students i) under the old assessment policy (n = 648) and ii) under the new assessment policy (n = 529) completed the Motivated Strategies for Learning Questionnaire on SRL, and additional items on participation. Year-1 performance was operationalised as students' average Year-1 course examination grades. MANOVA and structural equation modelling were used for analyses.

Results: Generally, students under the new assessment policy showed significantly higher Year-1 performance, SRL and participation, compared with students under the old assessment policy. The relationships between Year-1 performance, SRL and participation were similar across assessment policies.

Conclusions: This study indicates that the higher academic performance under a compensatory higher stakes, higher performance standard assessment policy, results from higher SRL and participation, but not from altered relationships between SRL, participation and performance. In sum, assessment policies have the potential to optimise student learning and performance.

Keywords: assessment policy; academic performance; self-regulated learning; participation

Introduction

Optimising student learning and academic performance is a continuous challenge for medical schools. Since several studies have shown that 'assessment drives learning' (Al-Kadri et al., 2012; Cilliers et al., 2011; Heeneman et al., 2015), modifying the assessment policy may be an efficacious way to improve student learning and to enhance academic performance (e.g. average grades). For instance, there is empirical evidence that performance is superior on tests with higher stakes (Cole & Osterlind, 2008; Sundre & Kitsantas, 2004; Wolf & Smith, 1995) or higher performance standards (Elikai & Schuhmann, 2010; Johnson & Beck, 1988). Another line of research has shown that self-regulated learning (SRL; Pintrich & de Groot, 1990; Zimmerman, 1990) and participation in scheduled learning activities (Credé et al., 2010; Schmidt et al., 2009) are key predictors of academic performance, and reported on their joint contribution (Stegers-Jager et al., 2012). However, it is not known how assessment policies affect SRL, participation and performance for medical students. This study filled this gap by investigating whether average SRL, participation in scheduled learning activities and academic performance differ under two assessment policies, which vary in terms of stakes and performance standards. There is also a lack of research on how SRL and participation relate to academic performance under different assessment policies. As a starting point, we used a tested and cross-validated integrated model of SRL, participation and Year-1 medical student performance that was developed under a conjunctive, lower-stakes, lower performance standard assessment policy (Stegers-Jager et al., 2012). We tested whether this model could be cross-validated in a new sample of students who were subjected to a compensatory, higher stakes, higher performance standard assessment policy.

SRL, Participation and Academic Performance

Self-regulated learners i) are able to control their own effort and motivation; ii) reflect on their learning process and adapt this process when necessary; and iii) use proper behavioural strategies for learning, for instance summarising the literature (Pintrich & De Groot, 1990).

There is strong empirical evidence for the association of SRL with academic performance (Credé & Phillips, 2011; Kitsantas, 2002; Pintrich et al., 1993; Pintrich & De Groot, 1990; Richardson et al., 2012; Ross et al., 2003; Stegers-Jager et al., 2012; Sundre & Kitsantas, 2004). For instance, higher levels of several motivational constructs, such as intrinsic goals, self-efficacy and task value, have been shown to be associated with improved academic performance (Credé & Phillips, 2011; Pintrich et al., 1993; Pintrich & De Groot, 1990; Richardson et al., 2012; Stegers-Jager et al., 2012). The same holds for

learning strategies such as metacognitive self-regulation, elaboration, organisation, time management and effort regulation (Credé & Phillips, 2011; Pintrich et al., 1993; Richardson et al., 2012; Stegers-Jager et al., 2012). Composite scores of SRL are positively associated to academic performance as well (Kitsantas, 2002; Ross et al., 2003; Sundre & Kitsantas, 2004).

In addition to SRL, participation in scheduled learning activities is another important predictor of academic performance (Credé et al., 2010; Schmidt et al., 2009; Van den Berg & Hofman, 2005). Students' physical presence at lectures or other modes of instruction is a crucial predictor of higher academic performance (Credé et al., 2010). More individual study time also predicts higher academic performance (Schmidt et al., 2009; Van den Berg & Hofman, 2005).

A study by Stegers-Jager, Cohen-Schotanus and Themmen (2012) showed the joint contribution of SRL and participation towards academic performance. SRL was operationalised as 'motivational beliefs' and 'learning strategies' and measured with the Motivated Strategies for Learning questionnaire, which we also used in the current study (Stegers-Jager et al., 2012). 'Motivational beliefs' consisted of 'value' and 'self-efficacy', whereas 'deep learning strategies' and 'resource management' were indicators of 'learning strategies'. Positive associations between these components of SRL, participation in scheduled learning activities and academic performance were found, which indicated that higher SRL is related to higher participation and higher academic performance (Stegers-Jager et al., 2012). In addition, deep learning strategies showed a weaker but statistically significant negative direct link to average grade. In other words, although deep learning is positively associated to academic performance through resource management and participation, when controlling for this positive pathway, there is a negative association between deep learning and academic performance. In sum, previous research has shown that it is valuable to consider the joint contribution of SRL and participation towards academic performance.

The Role of Assessment Policies

Several studies have shown that raising the stakes (i.e. higher consequences of performance) is associated with superior academic performance (Cole & Osterlind, 2008; Sundre & Kitsantas, 2004; Wolf & Smith, 1995) and increased motivation (Knekta, 2017; Simzar et al., 2015; Sundre & Kitsantas, 2004; Wolf & Smith, 1995). Higher performance standards (i.e. higher demands in order to pass) have also been associated with increased academic performance (Elikai & Schuhmann, 2010; Johnson & Beck, 1988). The available research on the interrelationships between SRL and academic performance shows that when the stakes are raised, motivation becomes

less predictive of performance in both high-school (Simzar et al., 2015) and college students (Sundre & Kitsantas, 2004; Sungur, 2007). By contrast, metacognition, as well as overall measures of learning strategies, are more important predictors of performance when the stakes are higher (Sundre & Kitsantas, 2004; Sungur, 2007).

However, none of these investigations focused on medical students, or included participation. In addition, studies investigating the effects of higher stakes (Cole & Osterlind, 2008; Knekta, 2017; Simzar et al., 2015; Sundre & Kitsantas, 2004; Sungur, 2007; Wolf & Smith, 1995), compared tests with no consequences (e.g. test does not count as part of the grade) to tests with consequences (e.g. test counts as part of the grade). In this study, we compared tests with consequences (e.g. students need to obtain all Year-1 credits within two years) to tests with even higher consequences (e.g. students need to obtain all Year-1 credits within one year).

The Current Research

In this study, we investigated the effect of assessment policies on SRL, participation in learning activities and academic performance of Year-1 medical students. Firstly, we compared the average scores on SRL, participation in learning activities and academic performance of student cohorts in the two assessment policies. We hypothesised that motivational beliefs and academic performance would be superior under higher stakes and higher performance standards. Based on the available literature, we were not able to formulate any hypotheses on learning strategies and participation.

Secondly, we examined whether the relationships between SRL, participation in learning activities and academic performance were similar under different assessment policies. Therefore, we tested whether the model that was developed by Stegers-Jager et al. (2012), was invariant for students under both assessment policies. In the case of higher stakes and higher performance standards, we expected that motivational beliefs would show weaker relationships with academic performance and that learning strategies would show stronger relationships with academic performance, compared with the lower stakes and lower performance standard assessment policy.

Method

Context

Both the initial study by Stegers-Jager et al. (2012) and the current study were performed with Year-1 students at the Erasmus MC Medical School, Rotterdam (the Netherlands). The curriculum consists of a 3-year Bachelor's programme, followed by

a 3-year Master's degree course. Year 1 of the bachelor programme consists of three thematic blocks and nine written exams. These exams are graded on a 10-point scale (1 = poor, to 10 = perfect) and consist of both closed and open-ended questions.

There are four types of learning activities: (i) large-group learning (lectures and patient demonstrations; 8 hours a week), (ii) small-group learning (skills training and tutorials; 8 hours a week), (iii) guided individual study (study assignments; 16 hours a week) and (iv) unguided individual study (8 hours a week). The small-group learning is compulsory for approximately a quarter of the meetings, the other learning activities are voluntary.

The only major curriculum alteration over the past years was the change in the assessment policy in 2014. The courses and the content of the curriculum have remained stable. The change to the assessment policy was made with the intention to accelerate academic progress of Year-1 students (Stegers-Jager & Themmen, 2015). In the previous lower stakes, lower performance standard, conjunctive *(old)* assessment policy (Stegers-Jager et al., 2011), students needed to obtain a sufficient grade (i.e. at least 5.5 out of 10) on each of nine examinations. Students were required to obtain 40 out of 60 possible Year-1 credits within the first year of enrolment in order to be allowed to proceed to the second year. After two years, all 60 Year-1 credits needed to be obtained to prevent academic dismissal. Students thus had three resit opportunities per examination, one in the first year and two more in the second year. In the new, higher stakes, higher performance standard, compensatory *(new)* assessment policy (Erasmus MC, 2014; Stegers-Jager & Themmen, 2015), obtainment of all 60 Year-1 credits within the first year of enrolment is compulsory in order for students to prevent academic dismissal. Therefore, per examination there is only one resit opportunity, resulting in higher stakes per individual examination, since the consequences of failing an assessment have risen. Also, an average grade of at least 6.0 is required; two grades of 5.0 – 5.49 are allowed under the condition that they are not obtained in the same thematic block. Thus, compensation is allowed, albeit minimal.

Hence, there are differences between the two assessment policies both in terms of the consequences of not obtaining all credits within the first year (i.e. the stakes), and in terms of the required grades in order to pass (i.e. the performance standards). It should be noted that another way for students to prevent academic dismissal once, is to drop-out before February, in which case students are allowed to re-enter Year 1 of the bachelor programme the next year. All the assessments are developed by an expert team in order to assure the quality and consistency of the assessments. Additionally, the Hofstee's method of standard setting is used for determining the pass/fail score per assessment (see Bandaranayake, 2008 for a detailed description

of the Hofstee method). The Hofstee method has been applied similarly under both assessment policies. Consequently, both the content of the assessments and the average pass/fail score have remained stable over the years. To balance out possible fluctuations in assessment characteristics that may have remained despite of these precautions, we used two cohorts for the old assessment policy, as well as two cohorts for the new policy.

Participants and Procedure

The participants of this study were Year-1 medical students, who enrolled in September 2008 and 2009 (old cohorts) or 2014 and 2015 (new cohorts). Each year, two months after enrolment, all Year-1 students were invited to voluntarily complete an online survey on SRL and participation in learning activities, which took 15-20 minutes. The students automatically received feedback on the basis of their SRL scores, providing information about their strengths and weaknesses, as well as recommendations for improvement. Students were informed about the study, in which they could voluntarily participate with guaranteed confidentiality. Because there was no plausible harm to participants in this study, the ethical committee of the Department of Psychology, Education and Child Studies of Erasmus University Rotterdam deemed further approval of a Medical Ethical Evaluation committee to be not required. Prior to analyses, all data were coded and saved without directly identifiable information.

Measures

Self-regulated learning (SRL) was measured with the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich et al., 1991), a thoroughly tested tool (Duncan & McKeachie, 2005), that is reliable and useful in predicting academic performance (Credé & Phillips, 2011; Pintrich et al., 1993), and proven appropriate in the medical context (Cook et al., 2011). In line with Stegers-Jager et al. (2012), we used a Dutch translation of the MSLQ (Blom & Severiens, 2008) for measuring motivational beliefs (subscales of intrinsic goal orientation, task value and academic self-efficacy), deep learning strategies (subscales elaboration, organisation and metacognitive self-regulation) and resource management (subscales time and study environment, and effort regulation), see Figure 1 for example items. Items are scored on a 7-point Likert scale (1 = *not at all true of me*, 7 = *very true of me*). Some items were minimally adapted to make them more suited for the specific medical school context, for instance by changing the word 'course' to 'theme' (c.f. Stegers-Jager et al., 2012).

Students were also asked to report on their participation in scheduled learning activities using three 5-point scale items about percentages of lecture attendance, skills training attendance and individual study assignment completion, see Figure 1.

MSLQ - Motivational beliefs

'In a class like this, I prefer course material that really challenges me so I can learn new things.' (Intrinsic goal orientation)

'I am very interested in the content area of this course.' (Task value)

'I expect to do well in this class.' (Self-efficacy for learning and performance)

MSLQ - Deep learning strategies

'When reading for this class, I try to relate the material to what I already know.' (Elaboration)

'When I study the readings for this course, I outline the material to help me organize my thoughts.' (Organization)

'If course materials are difficult to understand, I change the way I read the material.' (Metacognitive self-regulation)

MSLQ - Resource management

'I make good use of my study time for this course.' (Time and study environment management)

'I work hard to do well in this class even if I don't like what we are doing.' (Effort regulation)

Participation

'What percentage of the lectures did you attend?' (Lecture attendance)

'What percentage of the skills trainings did you attend?' (Skills training attendance)

'What percentage of the guided individual study assignments have you completed?' (Individual study assignment completion)

Figure 1. Example items from selected subscales of the Motivated Strategies for Learning Questionnaire, and participation items

Outcome Measure - Year-1 Performance

At the end of the academic year, we obtained students' grades for their first attempt at all nine Year-1 course examinations from the university student administration system. Next, we calculated Year-1 performance as the unweighted average of the grades for all students who earned at least seven grades, regardless of whether these grades were sufficient or not.

Statistical Analyses

After we screened for accuracy of data entry and missing values, and checked the study variables for normality, we calculated descriptive statistics and Pearson correlations, and Cronbach alphas for the subscales of the MSLQ. To examine differences in SRL, participation and performance between both assessment policies a MANOVA (Field, 2013) was performed. Analyses were performed using IBM SPSS Statistics for Windows Version 23.0 (2014). We checked Box's M to assess whether necessary assumptions were met. Next, we calculated Pillai's Trace for the overall model. In case of a significant outcome of the multivariate test, we performed univariate ANOVAs on the separate

dependent variables. We calculated *F*-values and Cohen's *d* (0.20 = small effect size; 0.50 = medium effect size; 0.80 = large effect size; Cohen, 1992) for the individual dependent variables.

We performed a multi-group analysis with structural equation modelling (SEM; Byrne, 2013), using AMOS 22.0 (Arbuckle, 2013), with students under the old assessment policy as the first group and students under the new policy as the second group. Structural equation modelling (SEM) combines factor analysis with regression, by creating latent constructs from observed scale scores, and then regressing these latent constructs on each other (Byrne, 2013). The goal of the multi-group SEM was to investigate whether there is structural invariance, meaning that the structural regression paths between the latent constructs are similar in both groups (Byrne, 2013; e.g. whether the regression path between the latent constructs deep learning and resource management is similar between groups). A necessary condition in order to assess structural invariance is measurement invariance. Measurement invariance means that the factor loadings, i.e. the connections between the latent constructs and their corresponding observed scale scores, are similar between groups (Meredith & Teresi, 2006). In other words, measurement invariance indicates whether the same construct is being measured across the specified groups, e.g. whether the observed scale scores for time management and effort regulation have similar loadings on the latent construct resource management in both groups.

In order to assess whether the factor loadings and structural paths were identical across groups, we added constraints in a stepwise manner. Firstly, to test measurement invariance, we constrained all factor loadings, error covariances and covariances to be equal across groups. Secondly, to test structural invariance, we constrained the structural paths to be equal, in addition to the constraints of the first step. Maximum likelihood estimation was used to estimate model parameters and a chi-square test to assess model fit was supplemented by the comparative fit index (CFI), the standardized root mean squared residual (SRMR), the root mean square error of approximation (RMSEA) and the Akaike Information Criterion (AIC). Since the chi-square test is strongly affected by sample size, the additional measures are necessary for evaluating model fit (Byrne, 2013). In general, the following results for these fit indices are considered good: a CFI ≥ .95, an SRMR ≤ .08 and an RMSEA ≤ .06 (Hu & Bentler, 1999).

Results

Respondents

The inclusion criteria for the study were that students completed the questionnaire, and attended at least seven out of nine possible assessments. For the 2008 and 2009 cohorts, 82% out of 817 Year-1 students completed the questionnaire, and 93% of the 817 students obtained at least 7 grades. In total 79% of the students met both inclusion criteria (n = 648, 35% male, M_{AGE} = 19.3 years, SD_{AGE} = 1.56 years). For the 2014 and 2015 cohorts, 79% out of 822 students completed the questionnaire, and 81% of the students obtained at least 7 grades as well. In total, 64% of the students met both inclusion criteria (n = 529, 33% male, M_{AGE} = 19.0 years, SD_{AGE} = 1.82 years). All respondents answered all items of the questionnaire.

Descriptive Statistics

The descriptive statistics for the study variables are presented in Table 1 and the correlations between the study variables are presented in Table 2. The Cronbach's alphas of the subscales for all four cohorts combined ranged from 0.61 to 0.87 (see Table 2). Overall, the correlations between the study variables were slightly lower under the new assessment policy, compared with the old policy.

Differences in SRL, Participation and Performance

The MANOVA with assessment policy as independent variable (IV) and students' scores on the eight separate subscales of the MSLQ, the three items for participation in scheduled learning activities and average grade as dependent variables (DVs), resulted in a highly significant Box's M (p < .001). Since Box's M test is sensitive to departures from normality, and the three participation variables were negatively skewed, we averaged the three participation variables into one participation variable and continued our analysis with this single participation variable. Thereafter, the assumptions for a MANOVA were met.

Table 1. Descriptives, p-values and effect sizes for the study variables (old cohorts [n = 648] and new cohorts [n = 529]).

	Variable	M_{old}	SD_{old}	M_{new}	SD_{new}	*p*	*d*
	Motivational beliefs						
1	Intrinsic goal orientation	5.74	0.73	5.79	0.72	n.s.	-
2	Task value	5.77	0.73	5.93	0.71	<.001	.22
3	Self-efficacy	4.89	0.84	5.08	0.80	<.001	.23
	Cognitive strategies						
4	Elaboration	4.85	0.87	4.86	0.90	n.s.	-
5	Organisation	4.66	1.16	4.89	1.23	.001	.19
6	Metacognition	4.27	0.80	4.60	0.83	<.001	.40
	Resource management						
7	Time management	4.63	1.04	4.91	1.01	<.001	.27
8	Effort regulation	4.91	1.06	5.33	0.97	<.001	.41
	Participation					.004	.17
9	Lecture attendance	4.69	0.67	4.78	0.62	-	-
10	Study assignments	4.10	1.15	4.06	1.18	-	-
11	Skills training attendance	4.58	0.67	4.84	0.47	-	-
	Year-1 performance						
12	Average grade	6.06	0.94	6.57	0.81	<.001	.57

M = mean; SD = standard deviation; n.s. = non-significant.

The multivariate test was significant for assessment policy, Pillai's Trace = 0.131, F (9, 1165) = 18.819, p < .001, indicating differences on the DVs between both assessment policies. Univariate analyses showed that students under the new assessment policy scored significantly higher on the measures task value (F (1, 1175) = 14.214, p < .001, d = 0.22), self-efficacy (F (1, 1175) = 15.676, p < .001, d = .23), organisation (F (1, 1175) = 10.655, p = .001, d = .19), metacognitive self-regulation (F (1, 1175) = 45.656, p < .001, d = .40), effort regulation (F (1, 1175) = 48.610, p < .001, d = .41), time management (F (1, 1175) = 21.154, p < .001, d = .27) and participation (F (1, 1175) = 8.554, p = .004, d = .17). Differences in average grade were also significant (F (1, 1175) = 99.554, p < .001, d = .57), with higher average grades for students under the new assessment policy. Hence, only differences in intrinsic goal orientation and elaboration were not statistically significant.

Table 2. Cronbach's Alphas (on the diagonal in bold, for all cohorts combined) and Pearson correlations for the study variables (old cohorts [n = 648] above diagonal, new cohorts [n = 529] below diagonal).

	Variable	n Items	1	2	3	4	5	6	7	8	9	10	11	12
	Motivational beliefs													
1	Intrinsic goal orientation	4	***0.61***	0.61*	0.51*	0.45*	0.27*	0.39*	0.27*	0.32*	0.13*	0.10†	0.10†	0.11*
2	Task value	6	0.56*	***0.85***	0.43*	0.41*	0.33*	0.39*	0.33*	0.43*	0.20*	0.14*	0.14*	0.13*
3	Self-efficacy	8	0.47*	0.39*	***0.87***	0.39*	0.17*	0.38*	0.36*	0.28*	0.06	0.05	0.03	0.18*
	Cognitive strategies													
4	Elaboration	6	0.44*	0.40*	0.33*	***0.68***	0.54*	0.60*	0.51*	0.42*	0.16*	0.22*	0.19*	0.17*
5	Organisation	4	0.20*	0.28*	0.08	0.53*	***0.74***	0.51*	0.42*	0.40*	0.21*	0.23*	0.13*	0.13*
6	Metacognition	10	0.34*	0.33*	0.29*	0.63*	0.48*	***0.77***	0.50*	0.46*	0.17*	0.21*	0.17*	0.17*
	Resource management													
7	Time management	5	0.24*	0.26*	0.30*	0.36*	0.34*	0.45*	***0.72***	0.69*	0.29*	0.52*	0.27*	0.32*
8	Effort regulation	4	0.31*	0.32*	0.20*	0.34*	0.29*	0.39*	0.60*	***0.74***	0.33*	0.51*	0.30*	0.35*
	Participation													
9	Lecture attendance	1	0.08	0.05	-0.04	0.08	0.07	0.04	0.17*	0.20*	-	0.33*	0.48*	0.33*
10	Study assignments	1	0.15*	0.14*	0.10†	0.18*	0.17*	0.19*	0.41*	0.40*	0.25*	-	0.30*	0.43*
11	Skills training attendance	1	0.15*	0.15*	0.07	0.05	0.03	0.06	0.13*	0.19*	0.45*	0.28*	-	0.29*
	Year-1 performance													
12	Average grade	-	0.01	0.02	0.09†	0.03	0.04	0.05	0.23*	0.25*	0.14*	0.38*	0.20*	-

* $p < .01$; † $p < .05$

Multi-Group Analysis of Structural Relationships

Results from the multi-group SEM indicated measurement invariance, since the CFI, RMSEA and SRMR were below the thresholds for proper model fit, see Table 3. Hence, the measurement models were equal between groups, indicating that the same factors were being measured in the old and new assessment policies. Additionally, the structural model (i.e. Model 3 vs. Model 2) was not significantly different across groups, indicating that the structural relationships were similar in the old and new assessment policies. The final model 3, with both measurement and structural invariance, had the smallest AIC (which is used to compare models), and showed good fit to the data: $\chi^2[108, n = 1177] = 354.835$, CFI = .947, SRMR = .048, RMSEA = .044, indicating that the model was invariant across assessment policies.

Consequently, there was a positive path from value through deep learning, resource management and participation to Year-1 performance, see Figure 2. There also was a negative direct relationship between deep learning and average grade, whereas self-efficacy showed a positive direct relation with average grade. The model explained 34% of the variance in average grades for students in the old assessment policy, and 32% of the variance for students in the new assessment policy.

Table 3. Goodness-of-fit statistics for tests of measurement and structural invariance across old and new assessment policies

Model description	Comparative model	χ2	*df*	CMIN/DF	Δdf	CFI	ΔCFI*	RMSEA	SRMR	AIC
1 Configural model; no equality constraints imposed	-	332.907	96	3.47	-	0.949	-	0.046	0.045	452.907
2 Measurement model, all factor loadings, error covariance and covariance constrained equal	2 versus 1	351.210	104	3.38	8	0.946	-0.003	0.045	0.048	455.210
3 Structural model, all factor loadings, error covariance, covariance and structural paths constrained equal	3 versus 2	354.835	108	3.29	4	0.947	0.001	0.044	0.048	450.835

CMIN/d.f. = chi-squared divided by the degrees of freedom; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardised root mean squared residual; AIC = Akaike information criterion.

* ΔCFI should be less than 0.01.

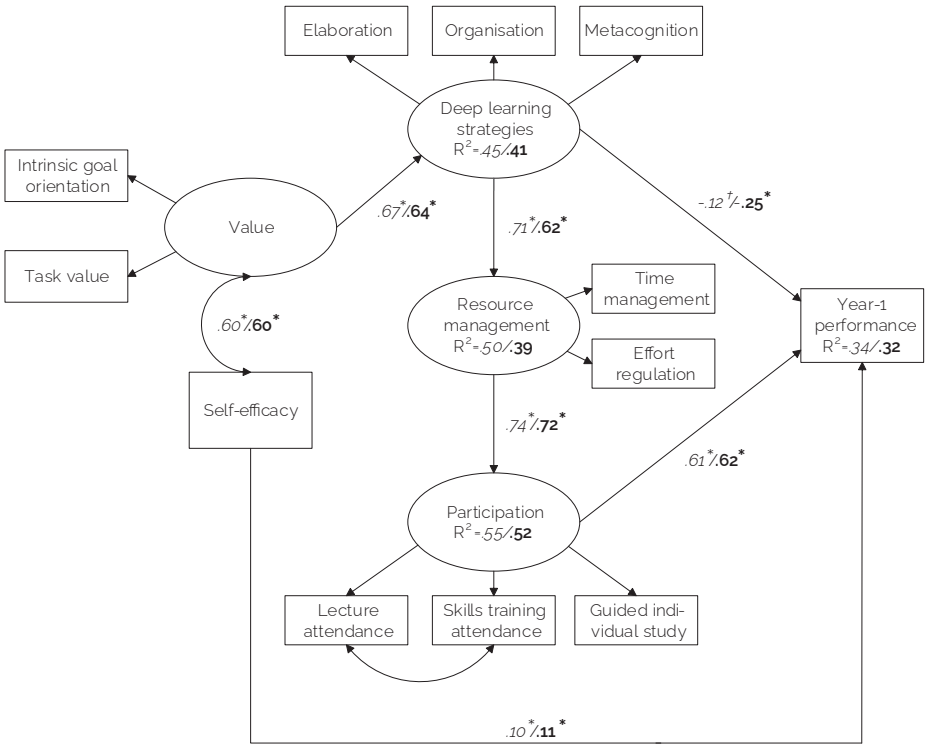


Figure 2. Multi-group model of Year-1 performance. Observed variables are represented by rectangles, latent constructs are represented by ovals. Results are *italic* for old group and **bold** for new group. Reported path values are standardised regression weights. * p < .001 and † p < .05, indicate whether the structural relationship per group is significant. R^2 is the proportion of variance accounted for that specific variable.

Discussion

This study showed that average grades were superior under a new assessment policy with higher stakes and higher performance standards, compared to an old policy with lower stakes and lower performance standards. Task value, self-efficacy, organisation, metacognition, effort regulation, time management and participation were significantly higher under the new policy, but intrinsic goal orientation and elaboration did not differ between both assessment policies. Additionally, the effect sizes for metacognition (*d* = .40), effort regulation (*d* = .41) and academic performance (*d* = .57) were substantial. The structural relationships between SRL, participation and academic performance were invariant, indicating that the relationships between SRL, participation and academic performance are similar in the two assessment policies. Thus, it seems

that the higher academic performance under the new assessment policy can be explained by increases in SRL and participation compared with the old assessment policy, although the ways in which SRL and participation affect performance are similar in both policies.

Higher Academic Performance, SRL and Participation

It is not surprising that academic performance improved after the stakes and performance standards were raised, because this is in line with previous findings (Cole & Osterlind, 2008; Elikai & Schuhmann, 2010; Johnson & Beck, 1988; Sundre & Kitsantas, 2004; Wolf & Smith, 1995). However, the magnitude of the increase in performance is striking, because it is identical to the raise in performance standards (i.e. half a point on a ten-point scale). This suggests that students are highly responsive to the minimal performance standards. It would therefore be interesting to further investigate the relation between demands of the assessment policy and academic performance.

Perhaps more surprising than the raise in performance, were the higher average scores for the new cohorts on the motivational construct task value, and the lack of a difference in intrinsic goal orientation. These results seem to contradict the notion that extrinsic motivators decrease, or have no influence on intrinsic motivation (Deci et al., 1999; Harlen & Crick, 2003). A possible explanation is that the number of extrinsic motivators, in this case examinations, has not been raised. Only the characteristics (i.e. the stakes and performance standards) of the extrinsic motivators were altered, and perhaps these characteristics now better match the students' performance level, as indicated by the higher self-efficacy of students in the new assessment policy. In other words, specific, difficult goals can be motivating, as long as the goals are deemed important and attainable (Locke & Latham, 2002).

Concerning self-regulated learning strategies and participation, we found higher scores on measures of deep learning (i.e. organisation and metacognitive self-regulation), resource management (i.e. time and study environment, and effort regulation) and participation for the new cohorts. An explanation that needs further examination is that when stakes and performance standards are raised, students increase the frequency of learning behaviours by which they expect to achieve success. The fact that elaboration did not increase significantly, would then indicate that students judge elaboration to be less important for achieving high grades. Overall, we found higher academic performance, SRL and participation in the new assessment policy, compared with the old policy.

Similar Relationships in Both Assessment Policies

The structural relationships between SRL, participation and academic performance in the model were comparable across both assessment policies, which indicates that SRL and participation were similarly related to academic performance under both policies. In short, higher value is associated with higher deep learning, which is related to better resource management, higher participation, and better academic performance. Self-efficacy shows a positive direct relation with academic performance. However, there is also a negative direct link from deep learning to academic performance. This may indicate that the Year-1 assessments do not reward deep learning, or alternatively that students need to combine deep learning with proper resource management and participation, in order to achieve academic success (Stegers-Jager et al., 2012).

Our results are somewhat surprising, because earlier research reported that when the stakes are raised, motivation shows weaker relationships with academic performance, and learning strategies and metacognitive strategies are more strongly related to academic performance (Sundre & Kitsantas, 2004; Sungur, 2007). However, we compared high stakes with even higher stakes, while these earlier studies compared low stakes with high stakes. In sum, it seems that SRL and participation are associated with academic performance in the same way under both assessment policies.

Limitations

The current study has several limitations that need to be addressed. First, we used correlational data, hence no firm causal conclusions can be drawn. Second, we used student responses on self-report questionnaires as measures of learning behaviours, which might be influenced by social desirability. Nonetheless, responding to the questionnaire was voluntary, confidential and the primary goal of the questionnaire was to aid students in self-reflection on the strengths and weaknesses of their study approach. Therefore, we do not expect answers to be shaped by social desirability. Third, we should note that the percentage of early drop-outs was higher in the new assessment policy (19%), compared with the old policy (7%). As it is likely that mainly students with low scores on our study variables dropped-out, this might partly explain the average differences between students under both assessment policies. However, we were able to check for differences on the basis of early dropouts who did complete the questionnaire, and still found comparable differences between the assessment policies on the study variables when they were included. Also, the standard deviations of our study variables were highly similar across both assessment policies, which contradicts the notion that only students with low scores on these variables dropped out. Moreover, this early selection is an effect the assessment policy may have,

discouraging some students from continuing their study while improving grades for those who stay (Lindo et al., 2010).

Another limitation of this study is the fact that we could only compare the results for the 2014 and 2015 cohorts with those for the cohorts from 2008 and 2009, because the MSLQ and participation questionnaire were not conducted in the years 2010 through 2013. Although no major alterations in the curriculum were made in this period, the selection procedure was changed in 2012. For the 2008/2009 cohorts 50% of students were admitted by weighted lottery and 50% were selected by a school-specific selection procedure, for an explanation of this procedure see Stegers-Jager et al. (2015). For the cohorts since 2012 these numbers were 20% and 80%, respectively. However, we do not expect this time gap or altered selection procedure to have influenced the results. First, research shows no differences in pre-university grade point average and Year-1 achievement between selected and lottery-admitted students (Urlings-Strop et al., 2009). Second, we were able to compare the average Year-1 grades for the 2012 and 2013 cohorts (i.e. the last cohorts under the old assessment policy) with those for the 2014 and 2015 cohorts, and found similar differences to those reported in the current study: the average grades for the 2012 and 2013 cohorts ($M = 6.09$, $sd = .97$) did not differ significantly from average grades for the 2008 and 2009 cohorts ($M = 6.06$, $sd = .94$), but were significantly different to those from the 2014 and 2015 cohorts ($M = 6.57$, $sd = .81$), $t(750) = -13.691$, $p < .001$. In sum, the significant change in academic performance did not seem to coincide with the change in selection procedure, but with the change in assessment policy.

Practical Implications and Suggestions for Further Research

An important practical implication of this study is that medical schools should be keenly aware of the influence their assessment policy has on student learning and academic performance. Although intrinsic motivation is important, external triggers may have a powerful additional effect on academic motivation (Hidi & Harackiewicz, 2000). Developing an assessment policy that boosts motivation might be an efficient way to challenge students to perform better. A meta-analysis showed that the goals that students have in terms of grades are one of the most important predictors of academic performance (Richardson et al., 2012). Although it seems likely that the stakes and performance standards will influence these grade goals, the connection of the assessment policy to students' grade goals and subsequent academic performance needs further exploration. Additionally, it would be interesting to separate the effects of higher stakes and the effects of higher performance standards on academic performance, in order to compare their relative contribution. Finally, in order to fully understand the effects of higher stakes and performance standards, an investigation

of the long-term consequences of these alterations is necessary. Many tests do not capture the full range of competencies and knowledge (UNESCO, 2016), or may negatively affect the motivation to learn, especially when the tests are high stakes (Harlen & Crick, 2003). Therefore, although we found higher task value and no differences in intrinsic goal orientation under the new assessment policy, it is important to monitor motivation for learning in the long term as well.

Conclusion

In conclusion, overall we found higher academic performance, SRL and participation for students under the new assessment policy compared to the old policy with lower stakes and lower performance standards, but no differences in intrinsic goal orientation and elaboration. Structural relationships between SRL, participation and performance were not different between the assessment policies, indicating that the relation of academic performance to these constructs is similar in both assessment policies. Thus, although SRL, participation and performance are higher under the new assessment policy, their associations remain the same. Hence, these results underscore the literature, showing that SRL and participation are important for explaining academic performance. In addition, it seems that this relation is relatively stable under different assessment policies and most importantly, that SRL, participation and performance can be improved by the design of assessment policies. In sum, characteristics of the assessment policy seem to play an important role in optimising student learning and academic performance.

CHAPTER 4

Assessment policies and academic performance within a single course: The role of motivation and self-regulation

This chapter has been published as:

Kickert, R., Meeuwisse, M., Stegers-Jager, K. M., Koppenol-Gonzalez, G. V., Arends, L. R., & Prinzie, P. (2019). Assessment policies and academic performance within a single course: The role of motivation and self-regulation. ***Assessment & Evaluation in Higher Education****, 44(8), 1177–1190. https://doi.org/10.1080/02602938.2019.1580674*

Abstract

Despite the frequently reported association of characteristics of assessment policies with academic performance, the mechanisms through which these policies affect performance are largely unknown. Therefore, the current research investigated performance, motivation and self-regulation for two groups of students following the same statistics course, but under two assessment policies: Education and Child Studies (ECS) students studied under an assessment policy with relatively higher stakes, a higher performance standard, and a lower resit standard, compared with Psychology students' policy. Results show similar initial performance, but more use of resits and higher final performance (post-resit) under the ECS policy compared with the Psychology policy. In terms of motivation and self-regulation, under the ECS policy significantly higher minimum grade goals, performance self-efficacy, task value, time and study environment management, and test anxiety were observed, but there were no significant differences in aimed grade goals, academic self-efficacy, and effort regulation. The relations of motivational and self-regulatory factors with academic performance were similar between both assessment policies. Thus, educators should be keenly aware of how characteristics of assessment policies are related to students' motivation, self-regulation and academic performance.

Keywords: assessment policy; academic performance; motivation; self-regulation

Introduction

When trying to encourage people to jump higher, a sensible option is to raise the bar. Analogously, the educational literature has consistently shown that assessment policies with higher standards are associated with better academic performance (Cole & Osterlind, 2008; Elikai & Schuhmann, 2010; Kickert et al., 2018). For instance, students perform better on knowledge assessments when a higher percentage correct answers is required to obtain the same grade (Elikai & Schuhmann, 2010; Johnson & Beck, 1988). However, little is known about the mechanisms underlying the association between assessment policies and academic performance.

In exploring the association between assessment policies and academic performance, we used motivation and self-regulation as a conceptual framework. Motivational and self-regulatory factors are among the most important correlates of academic performance (Richardson et al., 2012; Schneider & Preckel, 2017). In addition, motivation and self-regulation have the advantage of being relatively alterable, compared to more stable student factors such as conscientiousness (Poropat, 2009), high school grade point average (Sawyer, 2013) and socioeconomic status (Sirin, 2005). For instance, the motivational factor self-efficacy (Bandura, 1982), is "deemed to be modifiable at a relatively low cost" (Richardson, Abraham, and Bond 2012, 375). As such, motivational and self-regulatory factors are likely candidates to be affected by assessment policies.

However, earlier research on assessment policies (Cole & Osterlind, 2008; Elikai & Schuhmann, 2010) failed to include some of the most important motivational and self-regulatory factors that are associated with academic performance (Richardson et al., 2012), such as performance self-efficacy and effort regulation. Moreover, our recent study, which did take several of these factors into consideration, merely involved medical students (Kickert et al., 2018). Therefore, a first aim of this study was to replicate earlier findings (Elikai & Schuhmann, 2010; Johnson & Beck, 1988; Kickert et al., 2018) on the association of assessment policies with academic performance, in a real-life setting with higher education social science students. Secondly, we extended earlier research by incorporating the most important motivational and self-regulatory factors (Richardson et al., 2012; Schneider & Preckel, 2017) in our investigation of the relationship between assessment policies and academic performance.

Assessment Policies

In this study, we compared two assessment policies that differed in three respects: i) the stakes, ii) the performance standard, and iii) the resit standard. The *stakes* are

the consequence of failing one or more assessments. Higher stakes have repeatedly been associated with higher performance (Cole & Osterlind, 2008; Sundre & Kitsantas, 2004; Wolf & Smith, 1995).

The *performance standard* is determined by the minimum grade required on the assessment of a course, in order to obtain the course credits. Higher performance standards have been associated with higher academic performance in diverse course programmes such as accounting (Elikai & Schuhmann, 2010), psychology (Johnson & Beck, 1988) and medicine (Kickert et al., 2018).

The *resit standard* refers to the number of permitted resit opportunities. There are several reasons for limiting the number of resits that a student is allowed to take. Firstly, providing more resit opportunities has been associated with lower performance on the initial assessment, although more resit opportunities were not associated with differences in final grades (Grabe, 1994). Secondly, a resit is an extra opportunity to pass an assessment by chance (Yocarini et al., 2018). Thirdly, resits may offer an unfair advantage to the resit students, for instance due to additional practice opportunities for the resit students (Pell et al., 2009). However, promoting additional practice can also be viewed as a purpose of resits (Proud, 2015). Fourthly, there are concerns about negative effects resits may have on student learning, such as a reliance on second chances (Scott, 2012), or lower investment of study time (Nijenkamp et al., 2016).

Factors Associated with Academic Performance

In a meta-analysis, Richardson, Abraham, and Bond (2012) identified the motivational and self-regulatory factors most strongly associated with academic performance. We firstly examined the relationship between assessment policies and academic performance in terms of changes in these factors (e.g. students' motivation may be boosted by higher performance standards). Additionally, we examined changes in the relations between motivational and self-regulatory factors and performance (e.g., the association between students' motivation and performance may be moderated by the performance standards). We will first describe the four most important motivational factors that are associated with performance, and then turn to self-regulatory factors of academic performance.

Motivational Factors

The four motivational factors that show the strongest association with academic performance are academic self-efficacy, performance self-efficacy, grade goals and task value (Richardson et al., 2012). The first factor, academic self-efficacy, refers to students' general perceptions of their academic capability (Richardson et al., 2012).

Differences in academic self-efficacy have been associated with differences in stakes and in performance standards, but there is empirical evidence that the relation between academic self-efficacy and performance is similar under different assessment policies (Kickert et al., 2018).

The second motivational factor, performance self-efficacy, which is also referred to as grade expectation (Maskey, 2012), is the specific grade students expect to obtain (Vancouver & Kendall, 2006). Hence, whereas academic self-efficacy is a relatively general measure of expectations concerning successful learning and performance, performance self-efficacy is more specific, focusing on the expected grade. Although performance self-efficacy is the strongest predictor of academic performance (Richardson et al., 2012), to the best of our knowledge there is no research on performance self-efficacy under different assessment policies.

A similar gap in the literature exists concerning the third motivational factor, students' grade goals under different assessment policies. The grade goal is a student's level of aspired grade (Locke & Bryan, 1968). Good grades are a primary focus for most students (Gaultney & Cann, 2001). As the assessment policies determine which grades are sufficient to pass a course, these policies also partially determine what students consider to be a good grade. Therefore, student grade goals are likely to be related to the assessment policies.

The fourth motivational factor is task value, which refers to a student's self-motivation for and enjoyment of academic learning and tasks (Richardson et al., 2012). Previous research has shown higher task value under higher stakes and performance standards, and similar relations between task value and academic performance under different assessment policies (Kickert et al., 2018). These results can be explained by the fact that setting specific difficult goals can be motivating, as long as these goals are deemed attainable (Locke & Latham, 2002). However, there have been concerns about the impact of external motivators, such as assessment, on students' intrinsic motivation (Deci et al., 1999; Harlen & Crick, 2003). Therefore, a replication of earlier findings concerning task value under different assessment policies would be useful.

In terms of the magnitude of the associations (Cohen, 1992), performance self-efficacy showed a large correlation with academic performance; the correlation with academic performance was medium-sized for grade goals and academic self-efficacy, and small-sized for task value (Richardson et al., 2012). Performance self-efficacy and grade goals were not included in previous investigations of the consequences of differences in assessment policies. These two motivational factors are important

predictors of academic performance and are intuitively likely to be influenced by assessment policies. Therefore, – next to academic self-efficacy and task value – performance self-efficacy and grade goals are important factors to take into account in order to understand the relationship between assessment policies and academic performance.

Self-Regulatory Factors

In addition to motivational factors, self-regulatory factors are important to consider when investigating academic performance (Richardson et al., 2012). Self-regulation entails that students are "metacognitively, motivationally, and behaviourally active participants in their own learning process" (Zimmerman 1986, 308). A first self-regulatory factor, effort regulation, can be defined as persistence and effort when faced with academic challenges (Richardson et al., 2012). Given that most students will at some point in their academic career encounter subjects that they deem less interesting (Uttl et al., 2013) or even anxiety-provoking (Onwuegbuzie & Wilson, 2003) the ability to sustain attention and effort in the face of distractions or uninteresting tasks seems to be a key factor in achieving academic success (Komarraju & Nadler, 2013).

A second important self-regulatory factor is time and study environment management, which refers to the capacity to plan study time and activities (Richardson et al., 2012). Time and study environment management has been found to be associated with academic performance, independent of intellectual correlates of performance, such as Scholastic Aptitude Test scores (Britton & Tesser, 1991). Effort regulation and time and study environment management have been shown to be higher under higher stakes and performance standards, although the association of both factors with academic performance is similar under different assessment policies (Kickert et al., 2018).

A third self-regulatory factor is test anxiety, which is considered to be the affective component of self-regulated learning (Pintrich, 2004). Test anxiety is the experience of negative emotions during test-taking situations, and is negatively related to intrinsic motivation, effort regulation and academic performance (Pekrun et al., 2011). Test anxiety is especially salient during statistics courses (Onwuegbuzie & Wilson, 2003). As the current research took place during a statistics course, we included test anxiety in this study.

The correlation between effort regulation and academic performance is medium-sized, whereas time and study environment management, and test anxiety show a small-sized association with performance (Richardson et al., 2012). To the best of

our knowledge, test anxiety was not taken into account in previous research into consequences of altered assessment policies.

Research Questions & Hypotheses

The first research question (RQ1) was whether we could replicate the earlier reported finding that academic performance is superior under more difficult assessment policies (Cole & Osterlind, 2008; Elikai & Schuhmann, 2010; Kickert et al., 2018). In the current research, we hypothesised this difference in performance to be present as well (H1).

Furthermore, we extended prior research by investigating the relationship between assessment policies and academic performance (RQ2). We therefore compared the most important motivational and self-regulatory constructs (Richardson et al., 2012) under two assessment policies that differed in terms of the stakes, performance standard and resit standard (i.e. RQ2a). On the basis of earlier research (Kickert et al., 2018), our hypothesis was that academic self-efficacy, task value, effort regulation and time and study environment management are higher under more difficult assessment policies (H2a). The current study extended previous research by including performance self-efficacy, grade goals and test anxiety.

Finally, we investigated whether the associations of these motivational and self-regulatory factors with academic performance are different under different assessment policies (i.e. RQ2b). On the basis of earlier findings (Kickert et al., 2018), we hypothesised that the associations of motivation and self-regulation with academic performance are similar under different assessment policies (H2b).

Methods

Educational Context

The current study was performed in the Bachelor's (BA) programmes of Psychology as well as Education and Child Studies (ECS) at a large urban university in the Western part of the Netherlands. The first two years of both 3-year BA programmes consist of eight consecutive 5-week courses; the third year consists of three (ECS) or four (Psychology) 5-week courses, a minor and a thesis and/or internship. At the end of each course, there is a written knowledge assessment that is graded on a 10-point scale (1 = *poor*, to 10 = *perfect*).

In February and March 2017, students from both course programmes took the same statistics course 'Psychometrics, an introduction'. The course consisted of

nine mandatory small-group meetings, six optional large-group lectures, and was concluded with a multiple-choice knowledge assessment. Since students from both course programmes followed the same course, they received identical instructional activities, course materials and assessments. However, for Psychology students this statistics course is part of BA-2, whereas the same statistics course is a BA-3 course for ECS students. Since the BA- 2 assessment policy differs from the BA-3 policy for both programmes, the same course is covered by different assessment policies for the two BA programmes.

Assessment policies

Psychology

In the Psychology curriculum, students are allowed to enter BA-3 without passing BA-2 entirely, including the statistics course currently under study. Therefore, the *stakes* of this BA-2 assessment are relatively low. Nevertheless, Psychology students do need to pass their entire BA programme in order to start with the Master's programme. The BA-2 Psychology assessment policy is compensatory, in that students need to obtain a Grade Point Average (GPA) of 6.0 for the eight assessments. Grades below 4.0 are considered invalid, and not compensable by higher grades. Thus, the *performance standard* is 4.0 for individual 5-week courses, as long as the overall BA-2 GPA is at least 6.0. BA-2 Psychology students are allowed a maximum of two resits for the eight BA-2 knowledge assessments. All resits take place in July after the academic year has ended, there is a maximum of one resit per course, and the highest attained grade counts. As the number of resits is limited for Psychology students, the *resit standard* is relatively strict.

Education and Child Studies

BA-3 ECS students are required to have passed BA-2, and need to pass the entire BA programme in order to progress to the Master's programme. This means that if students fail at least one BA-3 course after the resit, this failure will result in one year of academic delay. Therefore, the *stakes* of the BA-3 ECS assessment are relatively high, compared to the stakes for the BA-2 Psychology assessment. The BA-3 ECS curriculum has a conjunctive assessment policy, which entails that students need to pass each separate assessment with a minimum grade of 5.5. Thus, for ECS students the *performance standard* is 5.5 for individual courses. ECS students are allowed to retake all three third-year assessments once in July after the academic year has ended, and the highest attained grade counts. Therefore, the *resit standard* is relatively lenient. In sum, compared to the Psychology assessment policy, in the ECS policy the stakes are higher, the performance standard is higher, but the resit standard is more lenient. Hence, two out of three characteristics of the assessment policy were more difficult

in the ECS policy. Therefore, we considered the ECS policy to be more difficult than the Psychology policy.

Procedure

Students who followed the 5-week course 'Psychometrics, an introduction', received a paper questionnaire at the start of the ninth and final small-group meeting of the course in March 2017, on the Tuesday of the fifth week. Completion of the questionnaire took 5-10 minutes and was completely voluntary. All students were informed about the study and active informed consent was given by all respondents. The course knowledge assessment took place on Thursday in week 5 and the resit took place approximately four months later, in July 2017.

Participants

Participants for this study were BA-2 Psychology students and BA-3 ECS students. In order to compare academic performance between the Psychology and ECS assessment policies (RQ1), we compared the grades between the entire cohorts (N_{Psy} = 219; N_{ECS} = 85). To investigate the relationship between assessment policies and academic performance (RQ2), we used a subsample of students who completed the questionnaire. Hence, the sample of Psychology students consisted of 150 students, i.e. a 68% response rate (M_{age} = 20.86, SD_{age} = 2.31, 20% male). The sample for ECS consisted of 51 students, i.e. a 60% response rate (M_{age} = 21.65, SD_{age} = 1.72, 8% male, 2% gender missing). Both the initial and final grades of the Psychology sample and the ECS sample were representative for the respective cohorts[1].

Materials

Motivational Factors

Participants completed two motivational subscales of a Dutch version of the Motivated Strategies for Learning Questionnaire (MSLQ; (Blom & Severiens, 2008; Pintrich et al., 1991): *Task Value* (e.g. 'I am very interested in the content area of this course.'; alpha = .85) and *Self-Efficacy for Learning and Performance* (e.g. 'I expect to do well in this class.'; alpha = .90). Items were scored on a 7-point Likert scale (1 = *not at all true of me*; 7 = *very true of me*). Subscale scores were computed by averaging the scores for the subscale items, under the condition of no more than one missing item per subscale. Some items were minimally adapted to adjust them to the specific educational context, for instance by changing the word 'class' to 'course'.

1 More details of the analyses are available upon request to the first author

In addition to the MSLQ-subscales, we posed two grade goal items and a performance self-efficacy item. These three items were each scored on a multiple-choice scale ranging from 1 to 10, with 0.5 point increments. *Grade goals* were measured through two items that were based on Locke and Bryan's (1968) original measurement of grade goals: i) 'Which grade are you aiming for on the course exam of this course?', and ii) 'What is the lowest grade you would be satisfied with for the course exam of this course?'. We termed the first item *aimed grade goal*, and the second item *minimum grade goal*. *Performance self-efficacy* was measured by asking 'Which grade do you expect to earn on the course exam of this course?'.

Self-Regulatory Factors

Participants also completed three self-regulatory subscales of the Dutch version of the MSLQ: *Effort Regulation* (e.g. 'I work hard to do well in this class even if I don't like what we are doing'; alpha = .73), *Time and Study Environment Management* (e.g. 'I make good use of my study time for this course'; alpha = .78) and *Test Anxiety* (e.g. 'When I take a test I think about the consequences of failing'; alpha = .83). The scoring, subscale computation, and adaptation of items was as described above for the motivational MSLQ subscales.

Other Variables

At the end of the questionnaire, students reported their age (in years) and gender (male/female).

Grades

Student grades were obtained through the course coordinator, who is one of the authors of the current study (GKG). Since the Psychology and ECS students were subjected to different resit standards, we used the grades after the initial assessment as well as after the resit. These grades were respectively termed *initial grades* and *final grades* (1 = *poor*, to 10 = *perfect*).

Statistical Analyses

Data Screening and Validity Checks

Before performing the analyses, we screened variables for missing values and normality, and checked relevant assumptions. One respondent only answered about half of the questionnaire and was removed from the sample. All MSLQ subscales, as well as course grades, were normally distributed. However, the two grade goal items were non-normally distributed, as many students indicated that their grade goals matched the performance standard.

Next, we performed two checks to strengthen the validity of our conclusions. These checks served to ensure that Psychology and ECS students were comparable in terms of performance and motivation in other courses. Firstly, we performed an independent t-test on our respondents' grades for a BA-1 statistics course. This BA-1 course was identical for Psychology and ECS students, including an identical assessment policy. In this BA-1 assessment policy, all 60 BA-1 credits needed to be obtained after one year to prevent academic dismissal (i.e. high stakes); the performance standard and resit standard were identical to the BA-2 Psychology assessment policy for both groups of students. Final grades for Psychology (n = 140; M = 5.97; SD = 1.18) and ECS respondents (n = 50; M = 6.27; SD = 1.49) were not statistically significantly different, $t(72.13) = -1.30$, $p = .199$.

Secondly, we checked whether grade goals and performance self-efficacy were similar for Psychology and ECS students in an earlier basic statistics course with the same assessment policy for both course programmes. This course was taken by the Psychology students of the current study, but a later cohort of ECS students. The students of these two course programmes did not differ significantly on any of the items ($p > .05$).

Main Analyses

In order to investigate possible differences in performance under different assessment policies (RQ1), we performed a t-test on the initial grades, and a t-test on the final grades. Additionally, we performed a chi-square test to assess whether different numbers of students took the resit under both policies.

To compare Psychology and ECS students' motivation and self-regulation (RQ2a), we performed a MANOVA with the two different assessment policies as the independent variable, and the five motivational (i.e. aimed grade goal, minimum grade goal, performance self-efficacy, academic self-efficacy, and task value) and three self-regulatory factors (i.e. effort regulation, time and study environment management, and test anxiety) as the dependent variables. We calculated Pillai's Trace for the overall model and in case of multivariate significance we performed univariate ANOVAs for the separate dependent variables. Also, we calculated Cohen's d (.20/.50/.80 = small/medium/large effect size; Cohen 1992) for the significant dependent variables.

We also investigated whether the association of the motivational and self-regulatory factors with academic performance was different under different assessment policies (RQ2b). To this end, we performed a five-step hierarchical forced entry multiple regression with initial grades as the dependent variable. We regressed on initial grades

instead of final grades, to minimise the interval between the measurement of the independent variables and the dependent variable. We included the motivational variables in the model before the self-regulatory variables, because motivation precedes self-regulation (Covington, 2000). In the first step we only included assessment policy. In the following models we cumulatively included: i) the five motivational variables, ii) the interactions between the assessment policy and the five motivational variables, iii) the three self-regulatory variables, iv) the interactions between the assessment policy and the three self-regulatory variables. For each of the five steps, we assessed whether the R^2-change was significant. The interaction variables added in step three and five are needed to answer RQ2b: significant interactions denote differences between assessment policies concerning the associations of the motivational and self-regulatory predictors with academic performance.

Results

Descriptive Statistics

Descriptive statistics, Cronbach's alphas and correlations for the study variables under both assessment policies are shown in Table 1. All study variables except test anxiety are significantly correlated to either initial or final grades, in both Psychology and ECS. Correlations between the study variables seem similar under both assessment policies. However, compared with Psychology the correlation between the study variables and final grades is lower in ECS. None of the Psychology and ECS students reported a minimum grade goal below the respective performance standards (4.0 for Psychology, 5.5 for ECS).

Table 1. Descriptives, Cronbach's Alphas (on the diagonal, for both assessment policies combined) and Pearson correlations for the study variables (psychology respondents [*n* = 150] above diagonal, Education and Child Studies respondents [*n* = 51] below diagonal)

	Variable	M_{PSY}	SD_{PSY}	M_{ECS}	SD_{ECS}	*n* Items	1	2	3	4	5	6	7	8	9	10
1	Aimed grade goal	6.83	1.29	6.71	1.00	1	-	.58*	.49*	.41*	.25*	.11	.13	-.20†	.39*	.38*
2	Minimum grade goal	5.24	1.06	5.74	0.50	1	.63*	-	.59*	.54*	.40*	-.06	.06	-.15	.35*	.36*
3	Performance Self-efficacy	5.40	1.25	5.87	0.95	1	.65*	.65*	-	.77*	.29*	.11	.13	-.37*	.41*	.44*
4	Academic Self-efficacy	3.95	0.98	4.18	0.96	8	.57*	.57*	.77*	***.90***	.46*	.05	.08	-.43*	.22*	.27*
5	Task value	4.39	1.06	4.81	1.03	6	.45*	.53*	.38*	.52*	***.85***	.08	.17†	-.02	.16†	.18†
6	Time management	4.43	0.99	4.97	0.90	8	.10	.13	.15	.17	.23	***.78***	.70*	.14	.16†	.15
7	Effort regulation	4.72	0.98	4.96	0.83	5	.05	.15	.10	-.01	.18	.60*	***.73***	.11	.23*	.23*
8	Test Anxiety	4.09	1.33	4.56	1.24	5	-.42*	-.46*	-.48*	-.46*	-.35†	-.08	-.04	***.83***	-.09	-.10
9	Course Grade – initial	5.75	1.30	5.66	1.40	-	.38*	.36*	.40*	.34†	.42*	.26	.40*	-.27	-	.95*
10	Course Grade - final	5.82	1.24	6.17	1.18	-	.23	.24	.18	.15	.38*	.34†	.35†	-.25	.77*	-

Note. M = mean; SD = standard deviation; $_{PSY}$ = psychology; $_{ECS}$ = education and child studies;

* $p < .01$

† $p < .05$

Differences in Performance (RQ1)

Concerning possible differences in academic performance between the ECS assessment policy (i.e. the combination of higher stakes, a higher performance standard, and a more lenient resit standard) and the Psychology assessment policy (RQ1), hypothesis 1 was partly confirmed: the initial grades of Psychology (M = 5.63, SD = 1.40) and ECS students (M = 5.69, SD = 1.36) did not differ significantly, $t(302) = -.32$, $p = .751$; however, the final grades were significantly higher for ECS students (M = 6.28, SD = 1.22) than for Psychology students (M = 5.72, SD = 1.34), $t(302) = -3.32$, $p = .001$, $d = .42$. ECS students took significantly more resits (36%) than Psychology students (5%), $\chi^2(1) = 50.86$, $p < .001$.

Differences in Motivation and Self-Regulation (RQ2a)

To assess possible differences in motivation and self-regulation between both assessment policies (RQ2a), we performed a MANOVA with the five motivational (i.e. aimed grade goal, minimum grade goal, performance self-efficacy, academic self-efficacy, and task value) and the three self-regulatory factors (i.e. effort regulation, time and study environment management, and test anxiety) as dependent variables. Although Box's M, as well as the Levene's tests for minimum grade goals and performance self-efficacy were significant, the largest variance was observed in the largest sample, i.e. Psychology. Therefore, we continued our analyses because our hypothesis testing would be conservative (Stevens, 2009). The multivariate test was significant for assessment policy, Pillai's Trace = .194, F (8, 192) = 5.76, $p < .001$, indicating differences on the dependent variables between both assessment policies. Univariate analyses indicated that compared with Psychology students, ECS students showed significantly higher minimum grade goals (F (1, 199) = 10.38, $p = .001$, $d = .52$), performance self-efficacy (F (1, 199) = 5.99, $p = .015$, $d = .40$), task value (F (1, 199) = 6.23, $p = .013$, $d = .40$) , time and study environment management (F (1, 199) = 11.95, $p = .001$, $d = .56$), and test anxiety (F (1, 199) = 4.76, $p = .030$, $d = .35$), see Table 1 for means and standard deviations for both assessment policies. Aimed grade goal, academic self-efficacy, and effort regulation did not differ significantly between the Psychology and ECS students. Thus, hypothesis 2a was partly confirmed.

Differences in Associations With Initial Performance (RQ2b)

As shown in Table 2, of the five steps of the regression analysis two steps showed statistically significant R^2_{change}: step two, in which the motivational variables were added, $R^2_{change} = .24$, $F(5,194) = 11.99$, $p < .001$; and step four, in which the self-regulatory variables were added, $R^2_{change} = .04$, $F(3,187) = 3.30$, $p = .022$. The steps in which the interaction variables were added did not show statistically significant R^2_{change}. This indicates that the association of motivational and self-regulatory factors with initial grades is similar

Table 2. Results of the five-step hierarchical multiple regression analyses, with initial grades as dependent variable, and the assessment policy, motivational and self-regulatory variables, as well as the interactions of motivational and self-regulatory factors with assessment policy as independent variables (N = 201).

Predictors	$B_{Model\ 1}$ (SE)	$B_{Model\ 2}$ (SE)	$B_{Model\ 3}$ (SE)	$B_{Model\ 4}$ (SE)	$B_{Model\ 5}$ (SE)	95% CI $_{Model\ 5}$	$r_{x\text{-initial grades.all}}$
Constant	5.75* (0.11)	1.87* (0.56)	2.23* (0.62)	1.41 (0.79)	1.59 (0.83)	[-0.05 , 3.23]	
Assessment policy	-0.09 (0.22)	-0.31 (0.20)	-1.49 (2.05)	-1.29 (2.04)	-2.26 (2.75)	[-7.68 , 3.17]	-.06
Aimed grade goal		0.23* (0.09)	0.22† (0.09)	0.20† (0.09)	0.20† (0.09)	[0.02 , 0.39]	.16
Minimum grade goal		0.05 (0.12)	0.09 (0.13)	0.12 (0.13)	0.12 (0.13)	[-0.14 , 0.38]	.07
Performance self-efficacy		0.48* (0.12)	0.51* (0.13)	0.46* (0.13)	0.47* (0.13)	[0.21 , 0.73]	.25
Academic self-efficacy		-0.35† (0.15)	-0.43† (0.17)	-0.41† (0.17)	-0.41† (0.18)	[-0.76 , -0.07]	-.17
Task value		0.18 (0.09)	0.11 (0.11)	0.07 (0.11)	0.08 (0.11)	[-0.14 , 0.29]	.05
AP * GG aim			-0.07 (0.26)	-0.02 (0.25)	0.00 (0.25)	[-0.50 , 0.50]	.00
AP * GG minimum			-0.10 (0.51)	-0.21 (0.51)	-0.27 (0.51)	[-1.28 , 0.74]	-.04
AP * P-SE			-0.07 (0.35)	-0.10 (0.34)	-0.18 (0.35)	[-0.86 , 0.50]	-.04
AP * A-SE			0.29 (0.34)	0.34 (0.34)	0.44 (0.35)	[-0.24 , 1.13]	.09
AP * TV			0.31 (0.23)	0.30 (0.23)	0.24 (0.24)	[-0.23 , 0.70]	.07
Time management				-0.04 (0.12)	-0.01 (0.14)	[-0.29 , 0.27]	.00
Effort regulation				0.31† (0.12)	0.22 (0.14)	[-0.06 , 0.49]	.12
Test anxiety				-0.03 (0.07)	-0.02 (0.08)	[-0.18 , 0.14]	-.02
AP * TM					-0.10 (0.27)	[-0.63 , 0.44]	-.03
AP * ER					0.43 (0.29)	[-0.14 , 1.00]	.11
AP * TA					-0.02 (0.18)	[-0.37 , 0.33]	-.01
R^2	.00	.24	.26	.29	.30		
F	.17	10.02*	5.87*	5.49*	4.67*		
R^2_{change}	-	.24	.02	.04	.01		
F_{change}	-	11.99*	0.91	3.30†	0.90		
Adjusted R^2	.00	.21	.21	.24	.24		

Note. CI = confidence interval; $r_{x\text{-initial grades.all}}$ = partial correlation between the variable and initial grade, corrected for all other variables in model 5; AP = Assessment policy; GG aim = aimed grade goal; GG minimum = minimum grade goal; P-SE = Performance self-efficacy; A-SE = Academic self-efficacy; TV = Task value; TM = Time management; ER = Effort regulation; TA = Test anxiety.

* $p < .01$

† $p < .05$

under both assessment policies, which confirms hypothesis 2b. Thus, the assessment policy does not moderate the association of motivation or self-regulation with initial grades. The variables that explained a significant proportion of variance in initial grades were aimed grade goal, performance self-efficacy, academic self-efficacy and effort regulation.

Conclusion and Discussion

The first research question was whether we would observe higher academic performance under the higher stakes, higher performance standard, and more lenient resit standard ECS assessment policy than under the Psychology assessment policy. There were no significant performance differences on the initial assessment. However, in line with our hypothesis, final performance was indeed higher in the more difficult ECS assessment policy. Thus, our first hypothesis was partly confirmed.

In our attempt to clarify the relationship between assessment policies and academic performance (RQ2), we first investigated mean differences in motivation and self-regulation between both policies (RQ2a). We found significantly higher minimum grade goals, performance self-efficacy, task value, time and study environment management, and test anxiety in the ECS policy, but no significant differences in aimed grade goals, academic self-efficacy and effort regulation between the assessment policies. Thus, hypothesis 2a is partly confirmed. Concerning the relations of motivation and self-regulation with academic performance (RQ2b), in line with hypothesis 2b we found no significant differences in these relations between both assessment policies.

Academic Performance

Although the higher final performance under the ECS assessment policy is in line with the literature (Cole & Osterlind, 2008; Elikai & Schuhmann, 2010; Kickert et al., 2018), the lack of a significant difference in initial performance is not. It seems that ECS students may have delayed their higher performance until the resit. Since the ECS students had a more lenient resit standard, these students had the guaranteed opportunity to retake the assessment, and thus had the option to postpone their effort until the resit. As ECS students took significantly more resits than Psychology students, our results may confirm concerns about the consequences of resits, such as a reliance on second chances (Scott, 2012), lower performance on the initial assessment (Grabe, 1994), and lower investment of effort for the initial assessment (Nijenkamp et al., 2016). However, an alternative explanation is that ECS students were more incentivised to attempt to

improve their grade in the resit, as these students performed under higher stakes, and a higher performance standard than Psychology students.

Motivational Factors

In terms of motivation, we observed higher performance self-efficacy for ECS students compared with Psychology students. A possible explanation for this finding may be that specific, difficult goals are motivating, as long as these goals are deemed attainable (Locke & Latham, 2002). However, there was no significant difference in academic self-efficacy between both assessment policies. Thus, although ECS students expected a higher grade, judgements of relatively general academic capability did not differ between both policies. Therefore, these findings are an indication that performance self-efficacy and academic self-efficacy are separate constructs. Compared to academic self-efficacy, performance self-efficacy seems more susceptible to differences in assessment policies.

Minimum grade goals were significantly higher under the ECS policy, but there were no differences concerning aimed grade goals. A possible explanation is that the performance standard only determines which grade students consider sufficient, but not which grade students consider good. This needs further exploration, as it has been previously asserted that students dichotomously view grades as either 'good' or 'bad' (Boatright-Horowitz & Arruda, 2013).

Lastly, task value was significantly higher for ECS students. Although this is in line with previous findings (Kickert et al., 2018), it is surprising in the light of the assertion that extrinsic motivators, such as assessments, damage intrinsic motivation (Deci et al., 1999; Harlen & Crick, 2003). However, we should note that the ECS students did not have more or different assessments, but only different standards. These standards were more difficult and thus perhaps more motivating.

Self-Regulatory Factors

In terms of self-regulation, for the ECS assessment policy we found significantly higher time and study environment management, as well as higher test anxiety, compared with the Psychology policy. Thus, given the higher stakes and higher performance standard in the ECS policy, ECS students may be more inclined to properly manage their time and study environment. However, the higher demands also seem to result in more test anxiety. Lastly, contrary to previous findings (Kickert et al., 2018), there were no significant differences in effort regulation between both policies. Possible explanations for this discrepancy are that the earlier work involved medical students, or that the sample size of the current investigation was insufficient to detect an effect.

In sum, more research is needed to draw firm conclusions about effort regulation under different assessment policies.

Differences in Associations With Performance

Our results showed similar relations of motivation and self-regulation with academic performance under both assessment policies, in line with previous findings (Kickert et al., 2018). Thus, the higher academic performance under the higher stakes, higher performance standard, lower resit standard assessment policy, seems to result from higher motivation and self-regulation, but not from different associations of motivation or self-regulation with performance.

We should note, that in our regression analysis the most important predictors of academic performance were performance self-efficacy, aimed grade goals, academic self-efficacy, and effort regulation. Although performance self-efficacy, academic self-efficacy, and effort regulation were higher in the ECS policy, only performance self-efficacy was significantly so. Thus, the assessment policy may not affect all the most important predictors of performance. For instance, although the minimum grade goal was related to the assessment policy, the aimed grade goal was not.

Limitations

The current study had several limitations that need to be addressed. Firstly, no causal conclusions can be drawn, as all data were observational. Besides different assessment policies, there were other differences between both groups, such as age and the attended course programme. However, to strengthen the validity of our conclusions, as reported in the methods we performed two checks that affirmed the groups' comparability in terms of performance and motivation in other courses. Secondly, the sample size for ECS may not have been large enough to obtain sufficient power (Field, 2013). Thus, research with larger samples is needed. Thirdly, given the current conjunction of differences in the stakes, performance standards and resit standards, it is not possible to draw conclusions on separate effects of these three characteristics of assessment policies.

Implications and Suggestions for Further Research

To the best of our knowledge, the current study was the first to include all the most important motivational and self-regulatory predictors of performance in an investigation of assessment policies. However, as the current study was performed in a statistics course in social sciences course programmes, future studies could investigate whether similar conclusions are drawn in other types of courses and/or course programmes. Additionally, it would be interesting to compare assessment

policies that only differ in one respect, in order to draw conclusions about the separate elements of the policies.

In order to better explain changes in academic performance due to changes in assessment policies, other measures of student learning could be investigated as well. For instance, it would be interesting to see how the quantity and quality of students' use of time are affected. Moreover, students' well-being and stress levels could be taken into account, in order to monitor possible negative impacts of assessment policies. Furthermore, although motivation may be higher in the short-term, this may not be the case in the long-term. Therefore, enduring effects of assessment policies on motivation need to be monitored as well.

Given the fact that performance self-efficacy and aimed grade goal are both one-item measures, it is promising that these two constructs explain significant variance in academic performance. Therefore, it could be worthwhile to further investigate these two motivational measures, for instance by researching what types of students exist in terms of these measures.

Although changes to stakes, performance standards and resit standards seem to be rare, these changes require relatively little effort. Given our findings, these efforts seem highly effective in terms of gains in motivation, self-regulation and academic performance. However, aimed grade goals, academic self-efficacy, and effort regulation did not differ significantly between both assessment policies. Hence, more research is needed on how these predictors of performance can be improved through educational interventions as well.

Conclusion

Students' academic performance, motivation and self-regulation are sensitive to characteristics of the assessment policy. This makes sense, as all students wish to obtain a diploma, and thus need to perform to the standards of the assessment policy. Therefore, educators should be aware of the influence that their standards and expectations have on students' academic performance: higher bars may lead to higher jumping.

CHAPTER 5

Grade goals and performance self-efficacy throughout the first academic year: A latent class approach

This chapter is submitted as:

Kickert, R., Meeuwisse, M., Jongerling, J., Stegers-Jager, K.M, Arends, L.R., & Prinzie, P. (submitted). Grade goals and performance self-efficacy throughout the first academic year: A latent class approach.

Abstract

Grade goals and performance self-efficacy are the two motivational factors most strongly associated with academic performance in higher education. Although the first academic year is crucial in higher education, little is known about how grade goals and performance self-efficacy develop during the first year, and whether this development is similar for all students. Therefore, our aim was to identify latent motivational classes based on students' grade goals and performance self-efficacy throughout the first academic year, and to interpret, characterise and validate these classes. For social science students from three different subsamples (total N = 587), a two-class model fitted the data best. Throughout the first academic year, both classes displayed relatively stable grade goals and performance self-efficacy, except for a drop regarding the statistics course. Class 1 students' grade goals and performance self-efficacy were close to the performance standard. Therefore, we labelled class 1 the *sufficient motivation* class (66% of students). Class 2 had higher average grade goals and performance self-efficacy on all eight courses. Therefore, we described class 2 as the *high motivation* class (34% of students). Students in the *high motivation* class had significantly higher scores on various course evaluation items (e.g. 'I consider the subject of this course interesting'), as well as higher grades than the *sufficient motivation* class. The results suggest that up to 66% of students are motivated to minimally pass assessments. Therefore, teachers should carefully determine the best passing standards, so that students are optimally encouraged to learn and perform.

Keywords: motivation, higher education, grade goals, performance self-efficacy, latent class analysis

Introduction

The two motivational factors that show the strongest association with academic performance in higher education are students' grade goals and performance self-efficacy (Richardson et al., 2012; Schneider & Preckel, 2017). Grade goals are the grades students want to obtain; performance self-efficacy refers to the grades students expect to obtain (Richardson et al., 2012). However, despite the comparatively sizeable associations with academic performance, a recent review of meta-analyses on achievement in higher education included only thirteen effect sizes on grade goals and four on performance self-efficacy, compared to sixty for both performance goal orientation as well as learning goal orientation (Schneider & Preckel, 2017). Thus, grade goals and performance self-efficacy have received relatively little attention from motivational researchers.

Although there have been investigations of how grade goals (Radosevich et al., 2004) and performance self-efficacy (Khachikian et al., 2011) develop throughout a single course, it is not yet known how these two motivational constructs develop throughout an academic year, or more specifically: the first year. The first academic year is the starting point of a student's academic career, and is often a challenging transition phase (Van Herpen, 2019). Additionally, in Dutch higher education, students' performance in the first year is used to determine whether students are allowed to continue their Bachelor's programme (Arnold, 2015). Due to the importance of students' first-year performance, students' motivation in the first year is pivotal. Therefore, we investigated students' grade goals and performance self-efficacy throughout the first academic year.

In recent years, person-centred approaches to data analysis are gaining ground in the field of student motivation (Conley, 2012; Linnenbrink-Garcia et al., 2018; Pastor et al., 2007). In a person-centred approach, students are grouped into latent motivational classes that represent types of students. The advantage of this approach is that different types of developmental patterns can be distinguished. Therefore, in this study we aimed to identify latent classes of students based on grade goals and performance self-efficacy throughout all courses of the first academic year.

Student Motivation

In the current investigation, we examined two motivational constructs: grade goals and performance self-efficacy. Although no specific motivational theory was mentioned in the first publication on the grade goal construct (Locke & Bryan, 1968), grade goals seem intuitively related with goal setting theory (Locke & Latham, 1990). Goal

setting theorists have shown that specific, difficult goals lead to higher performance than more general 'do your best' goals (Locke & Latham, 2002). The association between goals and performance is most pronounced when stakes and self-efficacy are high (Locke & Latham, 2002). Consequently, high grade goals should lead to the best performance, especially in a high-stakes educational context. Indeed, studies have consistently shown that grade goals are positively associated with academic performance (Kickert et al., 2019; Locke & Bryan, 1968; Vancouver & Kendall, 2006; Wood & Locke, 1987). The only longitudinal investigation of grade goals we are aware of, showed that students tend to lower their grade goals throughout a ten-week course, and on average performed below their goals (Radosevich et al., 2004). Additionally, it was concluded that students lowered their grade goals in case of disappointing performance. However, this investigation only spanned a single course, instead of an entire academic year.

The only student factor that is more strongly associated with academic performance than grade goals, is performance self-efficacy (Richardson et al., 2012; Schneider & Preckel, 2017). Beliefs about competence play a role in all major motivational theories (Cook & Artino, 2016) and self-efficacy is a specific competence belief that has become fundamental to many motivational theorists (Bandura, 1982). In contrast to the more general concept of self-confidence, self-efficacy refers to a person's capability beliefs about performing a specific task. In an academic context, the term self-efficacy usually refers to the concept of academic self-efficacy. Academic self-efficacy denotes students' general academic capability beliefs, thus including students' self-efficacy for both learning and performance (Richardson et al., 2012). Compared to academic self-efficacy, performance self-efficacy is more specific, as it only concerns capability beliefs about assessment performance, i.e. grades (Richardson et al., 2012).

Previous longitudinal research has shown that most students' performance self-efficacy is optimistic at the start of a course, and more realistic towards the end (Guillaume & Khachikian, 2011; Hossain & Tsigaris, 2015; Khachikian et al., 2011). In other words, performance self-efficacy decreases over the span of a course, and becomes more accurate. The overall decrease may be particularly due to low-performing students' decreasing performance self-efficacy, as high performers' performance self-efficacy is stable (Khachikian et al., 2011). High-performing students' expectations are more accurate and show more accuracy improvement over multiple assessments than low-performing students' expectations (Hacker et al., 2000). Students with high performance self-efficacy also devote more time to their studies (Guillaume & Khachikian, 2011). However, as for grade goals, the only longitudinal investigations of performance self-efficacy we were able to find, spanned a single course (Guillaume & Khachikian, 2011;

Hacker et al., 2000; Hossain & Tsigaris, 2015; Khachikian et al., 2011). Thus, there are gaps in the literature concerning grade goals and performance self-efficacy throughout an entire academic year.

These gaps in the literature are remarkable, given that grade goals and performance self-efficacy are the two motivational factors that show the strongest association with academic performance (Schneider & Preckel, 2017). Therefore, gaining a more profound understanding of these two motivational constructs can help to further our understanding of academic success. Given the gaps in the literature, the relatively strong association of grade goals and performance self-efficacy with performance, and the importance of the first year, we investigated grade goals and performance self-efficacy throughout the first academic year. More specifically, we examined whether different types of students exist with regards to the development of grade goals and performance self-efficacy during the first academic year. To this end, we used a person-centred data analysis approach to motivation.

Person-Centred Approach

In many educational settings, performance standards determine the cut-off between passing and failing grades on assessments. In previous research, higher performance standards have been associated with higher grade goals (Kickert et al., 2019). In the traditional, variable-centred statistical approach (e.g., regression analysis), associations between variables are assumed to be homogeneous within the population; thus, an observed association between performance standards and grade goals is assumed to be similar for all students. Heterogeneity can be taken into account through interaction effects, but only in terms of interactions between observed variables. In that case, the association between performance standards and grade goals is assumed to be similar within all observed subgroups of students (Kickert et al., 2019).

However, some subgroups may not be observable, but are latent instead. Thus, different latent types of students may exist regarding the association between performance standards and grade goals. Students who aim to pass an assessment with minimal effort will likely hold a grade goal near or at the performance standard set by the faculty. Therefore, raising the performance standard may increase these students' grade goals, resulting in a positive association between the performance standard and grade goals for this first group. Conversely, other students may disregard the faculty's performance standard, because these students set their own performance standard. For these students, raising the standard will not affect their grade goal, thus this second group will show no association between the performance standard and grade goals. If the former subgroup is larger, the overall association between grade

goals and performance will be positive. However, this positive association will not be an accurate reflection of the entire population, due to the existence of latent types of students.

In order to take latent types of students into account, a person-centred approach has proven useful (Lambe & Bristow, 2011; Linnenbrink-Garcia et al., 2018; Pastor et al., 2007). In this approach, which encompasses analytical techniques such as Latent Class Analysis and Latent Profile Analysis, subjects are differentiated into various latent classes. Class membership is a latent variable that is derived from several observed variables, analogous to factor analysis. However, in the person-centred approach the latent variable is categorical instead of continuous. In other words, on the basis of several observed variables, subjects are categorised into different latent classes. Probabilities of belonging to each of the different classes are calculated for each subject. The goal is to form classes that are as homogeneous as possible within-class, and as heterogeneous as possible between-class; subjects within a class should be similar to each other, and dissimilar from subjects in other classes (Jung & Wickrama, 2008). After the classes have been identified, class membership can be associated with other variables of interest to gain insight into the nature of the classes (Lambe & Bristow, 2011). The associations between classes and other variables can also serve to establish criterion validity of the classes (Cronbach & Meehl, 1955).

The person-centred approach has already yielded valuable insights in the educational sciences. For instance, Lambe and Bristow (2011) showed that three latent classes based on prior achievement and admission interview scores could be used to predict medical student performance. In the motivational domain, studies have identified four to seven latent profiles on the basis of students' achievement goal orientation (Pastor et al., 2007), or based on a combination of competence beliefs, task value and achievement goal orientations (Conley, 2012; Linnenbrink-Garcia et al., 2018). In sum, the person-centred approach can be considered a valuable addition to educational research. Nevertheless, to the best of our knowledge, no previous studies have used latent class analysis to investigate heterogeneity in students' grade goals and performance self-efficacy during the first academic year.

Research Questions

Our first research question (RQ1) was: Which latent classes of students exist in terms of the development of grade goals and performance self-efficacy throughout the first academic year? Our aim was to elucidate how students shape their motivation around the performance standard throughout the first academic year, and which different types of students exist concerning this shaping behaviour. As we had no empirical

basis on which to ground hypotheses concerning the content and number of classes, the analyses were explorative.

Our second aim was to further characterise and validate the classes. Therefore, the second research question (RQ2) was: How does latent class-membership, based on grade goals and performance self-efficacy, relate with students' course evaluations and academic performance? As we did not know the nature of the latent classes, we could not form a priori hypotheses concerning RQ2.

Methods

Participants and Context

Participants for the current study were first-year Bachelor's (BA1) students at a large urban university in the Netherlands, who started in 2016 and completed at least one course evaluation at the end of one of the eight consecutive first-year courses. There were three subsamples of participants: psychology students from a Dutch track (n = 349), psychology students from an International track (n = 136), and Dutch education and child studies (ECS) students (n = 102; ECS has no international track). Psychology and ECS both have problem-based curricula. Thus, the majority of lessons are tutorial groups in which students discuss the literature with each other under the guidance of a tutor (Schmidt, 1994).

The international psychology students and the Dutch psychology students received the exact same lectures and instructional materials, all in English. However, there were language differences between the two subsamples in the tutorial group discussions and assessments of all courses, as well as in the lectures for the third course, Statistics 1; these were in Dutch for the Dutch students. The third (Statistics 1) and fifth (Developmental Psychology) courses were the only courses that are identical for psychology students and ECS students. Besides these two courses, ECS students have a different curriculum, although certain courses show substantive overlap with the psychology courses.

The first years of the psychology and ECS Bachelor's programmes consist of eight five-week block periods that comprise both a skills course and a knowledge course. Per week, students are expected to have a 40-hour workload. There are 10-12 contact hours per week: two tutorial groups, one lecture, and one skills training. Students are expected to perform 20 hours of individual study in preparation for the tutorial groups per week, and 5-10 hours in preparing for the skills training.

In the fifth and final week of each course, students take both a skills assessment and a knowledge assessment. Skills assessments take various forms, such as presentations or essays. The knowledge assessments are always written examinations that are all in multiple choice format for psychology, and mostly in multiple choice format for ECS. The knowledge assessments take place on the Thursday or Friday of the final week of a course. All grades are given on a ten-point scale (1 = lowest, through 10 = highest).

In both psychology and ECS, the eight BA1 skills courses form a 20-credits skills cluster, and the eight BA1 knowledge courses form a 40-credits knowledge cluster. The performance standard for both clusters entails that the respective GPA needs to be 6.0 or higher, and only grades equal to or higher than 4.0 are considered valid. Per cluster, a maximum of two resits is allowed, and all resits take place after the eighth block period. At the end of BA1, students are required to have obtained all 60 first-year credits in order to avoid academic dismissal and continue their course programme. Thus, students need to pass both clusters within one year. Consequently, the assessments are high stakes. In the current research we focused on the knowledge courses and assessments.

In order to subscribe for the knowledge assessments, students filled in a course evaluation in the week before the knowledge assessment, through the course website. Thus, most students completed eight evaluations. These evaluations were in English for all psychology students, and in Dutch for ECS students. The evaluations mostly consist of multiple-choice items about various aspects of the respective course, such as the quality of the lectures and group discussions, as well as students' own effort during the course. In addition, for one academic year (2016-2017), four motivational items were included in the evaluation, both for the current research (see below) and more broadly to evaluate the motivational consequences of the assessment policy.

Data from the course evaluations, as well as grades for the knowledge assessments, were obtained from the Erasmus Education Research Database in an anonymised form. In addition to the student evaluation data and grades, we did not obtain any further participant information, such as gender and age, as these data could potentially be used to identify individuals. Consequently, we obtained data that were not traceable to individual students. As this research would not reasonably be assumed to create distress or harm and involves the study of normal educational practices, we did not need to obtain informed consent (American Psychological Association, 2017).

Variables

Latent Class Motivational Variables

We used four motivational items from each course evaluation to construct the latent classes: two grade goal items and two performance self-efficacy items. As there were eight courses and four variables per course, there were 8 x 4 = 32 items per student. The motivational items were based on Locke & Bryan's (1968) original study, and three of these items were used previously as well in related research (Kickert et al., 2019). All these items only concerned the knowledge courses. Students' *grade goals* were assessed through 'Which grade are you aiming for on the course exam of this course?' and 'What is the lowest grade you would be satisfied with for the course exam of this course?'. We termed these items *aimed grade goal* and *minimum grade goal*, respectively. Students' *performance self-efficacy* was assessed through 'Which grade do you expect to earn on the course exam of this course?' and 'At the end of this academic year, an average grade of all 8 course exams will be calculated. What minimum average grade do you expect to earn for all 8 course exams?'. We termed these items *course performance self-efficacy* and *year performance self-efficacy*, respectively. The grade goal and performance self-efficacy items were all measured on a ten-point scale, corresponding to the grading scale. However, although grades were awarded with one decimal, only integer values were possible for the motivational items.

Outcome Variables

Student evaluation items. Besides the motivational items, we obtained several additional student evaluation items. For each of these items used as outcome variables (and thus not for the motivational variables used to form the latent classes), we averaged the scores over the eight courses. Use of time was measured by asking 'How much time on the average did you spent each week on independent study? (Fill in the answer in whole hours)'. Students' interest was measured by asking 'I consider the subject of this course interesting' (5-point scale, ranging from 1 strongly disagree to 5 strongly agree). Perceived relevance was measured by asking 'The relevance of this course within the context of my study was clear to me' (5-point scale, ranging from 1 strongly disagree to 5 strongly agree). Course rating was measured by asking 'If you had to mark this course program on a scale from 1 to 10, what mark would you assign to this course?'. Amount of learning was measured by asking 'I have learned a lot during this course' (5-point scale, ranging from 1 strongly disagree to 5 strongly agree).

Academic performance. We used the average grade for the eight knowledge assessments as a measure of academic performance. We excluded resit grades as these were only obtained after the eighth block period, up to 9 months after

the first evaluations were completed. Thus, we only used the grades on the initial assessments.

Statistical Analyses

First, we screened the data and checked the study variables for normality. In addition, per course we checked the descriptive statistics and bivariate correlations between the four motivational variables, as well as with course grade. The screening and checks were performed in IBM SPSS Statistics for Windows, Version 24.0 (2016).

Subsequently, to answer RQ1 we determined the number of latent classes in the pattern of scores on the four motivational variables in all eight courses (i.e. 4 x 8 = 32 class indicators). We used mixture modelling in Mplus version 7.4 (Muthén & Muthén, 1998-2012) with full information maximum likelihood and robust standard errors (i.e. MLR-estimator).

We started with a one-class model, and for each consecutive model we checked whether the addition of an extra class improved the fit significantly, compared to the previous model. The model fit improvement was evaluated using both the Vuong Lo-Mendell-Rubin (VLMR) and Lo-Mendell-Rubin (LMR) likelihood ratio tests. If these tests are significant ($p < .05$), this denotes a significant improvement of model fit by the addition of the extra latent class. Therefore, the best fitting model should be a significant improvement of the model with one fewer latent class, and should not be significantly improved by the addition of another latent class. We stopped adding more classes after two non-significant improvements. For instance, if the three-class model was a significant improvement over the two-class model, but the four- and five-class models were non-significant, a three-class model was chosen.

In addition to the likelihood ratio tests, for each model we evaluated the following criteria (Jung & Wickrama, 2008): 1) The percentages of students per class should not be too low, for instance at least 1% of the total count per class; 2) Entropy, which compares the within-class difference to the between-class difference and thus indicates the ability of the model to provide well separated clusters (Celeux & Soromenho, 1996), should be as close as possible to the maximum value of 1.0; 3) The average latent class probabilities, i.e. the average probability of belonging to a certain class for all students being assigned to that class, should be as close to the maximum value of 1.0 as possible. Although models can be compared on the basis of entropy and average latent class probabilities, no cut-off values are available.

After choosing the best fitting model, we assessed whether the model was invariant across the three different subsamples of students (Psy-Dutch/Psy-International/ECS), in order to assure that the model holds for each subsample. Therefore, we ran a multi-group analysis in which means on the 32 indicator variables were not only allowed to vary across latent classes, but also across the different subsamples. We compared this model with freely estimated indicator means to a model with fixed indicator means. In addition, we performed a multi-group analysis on the probabilities for students to belong to each class, to assess whether the distribution over the classes was similar in the three subsamples. We compared this model with freely estimated class probabilities to the model with fixed class probabilities. For both multi-group analyses, we calculated the Satorra-Bentler χ^2 difference, as it corrects for non-normality (Satorra, 2000). If the model is invariant, allowing variance over the three subsamples should not significantly improve model fit.

Finally, in order to further characterise the classes (RQ2), we performed six t-tests with class membership as the independent variable, and use of time, interest, relevance, course rating, amount of learning and average grades as the dependent variables. To compensate for the number of performed tests, we used a significance level of $p < .01$.

Results

Descriptive statistics

Data screening revealed that some students had suspicious response patterns on some evaluations (e.g. same answer on each item). Therefore, we deleted all evaluation data from one student, and segments of evaluation data from four students, before performing our analyses. In addition, one student who took one ECS course and all psychology courses, was removed from the ECS data.

The descriptive statistics per subsample of students for the grade goals and performance self-efficacy items, as well as grades, are presented in Table 1. Generally, within each subsample the means and standard deviations are comparable over the eight courses, with the exception of course 3, which is the only first-year statistics course for all three subsamples. Bivariate correlations between the grade goal and self-efficacy items within each course were always significant positive ($p < .001$), see Table 2. The correlations of the grade goals and performance self-efficacy with grades were always significant positive, except in the first course, and to a lesser extent in the fifth course.

Table 1. Means and standard deviations for the motivational variables and grades, for all three subsamples, per course.

	Course	$N_{evaluation}$ / N_{grade}	GG_{aim} Mean (sd)	$GG_{minimum}$ Mean (sd)	PSE_{course} Mean (sd)	PSE_{year} Mean (sd)	Grade Mean (sd)
Psy_{Dutch}	*1*	335 / 335	7.90 (0.88)	6.20 (0.60)	6.93 (0.81)	6.72 (0.64)	6.37 (1.16)
	2	308 / 310	7.45 (0.82)	6.04 (0.58)	6.54 (0.75)	6.54 (0.57)	6.91 (1.28)
	3	305 / 305	6.74 (1.23)	5.43 (1.07)	5.78 (1.23)	6.62 (0.68)	6.01 (1.46)
	4	293 / 293	7.41 (0.95)	6.13 (0.80)	6.71 (0.90)	6.60 (0.64)	6.54 (1.44)
	5	275 / 277	7.33 (0.88)	6.16 (0.68)	6.60 (0.79)	6.62 (0.65)	6.15 (1.25)
	6	283 / 287	7.57 (0.91)	6.32 (0.75)	6.75 (0.81)	6.55 (0.66)	7.02 (1.08)
	7	279 / 277	7.38 (0.98)	6.37 (0.69)	6.78 (0.77)	6.60 (0.63)	6.78 (1.05)
	8	275 / 274	7.20 (0.98)	6.18 (0.83)	6.72 (0.84)	6.63 (0.67)	7.04 (0.83)
Psy_{Int}	*1*	133 / 134	8.16 (0.96)	6.17 (0.66)	7.05 (0.91)	6.95 (0.80)	6.64 (1.40)
	2	122 / 124	7.64 (1.02)	6.16 (0.73)	6.85 (0.92)	6.66 (0.69)	7.43 (1.35)
	3	129 / 129	6.50 (1.46)	5.19 (1.24)	5.57 (1.51)	6.76 (0.77)	6.46 (1.68)
	4	123 / 124	7.80 (1.05)	6.35 (0.88)	6.89 (1.03)	6.93 (0.73)	7.15 (1.56)
	5	114 / 116	7.63 (0.94)	6.44 (0.96)	6.94 (1.08)	6.92 (0.87)	6.88 (1.26)
	6	119 / 120	7.82 (1.03)	6.56 (1.01)	7.02 (0.96)	6.97 (0.78)	7.52 (0.98)
	7	113 / 114	7.72 (1.00)	6.68 (0.95)	7.17 (0.91)	7.02 (0.76)	7.38 (1.14)
	8	116 / 117	7.68 (1.09)	6.63 (0.99)	7.09 (0.95)	6.96 (0.73)	7.23 (0.94)
ECS	*1*	98 / 98	7.66 (0.90)	5.87 (0.53)	6.62 (0.77)	6.37 (0.55)	6.42 (0.98)
	2	88 / 88	7.41 (0.88)	5.89 (0.47)	6.41 (0.58)	6.27 (0.47)	6.09 (1.16)
	3	88 / 86	6.50 (1.04)	5.23 (0.81)	5.61 (1.10)	6.26 (0.54)	6.11 (1.64)
	4	78 / 77	7.21 (0.95)	5.79 (0.49)	6.38 (0.63)	6.33 (0.55)	6.61 (1.15)
	5	73 / 72	7.07 (0.93)	5.82 (0.63)	6.25 (0.68)	6.30 (0.49)	6.70 (1.05)
	6	78 / 77	7.23 (0.95)	5.83 (0.65)	6.41 (0.65)	6.24 (0.56)	6.39 (1.08)
	7	75 / 75	7.09 (0.98)	5.81 (0.71)	6.41 (0.72)	6.37 (0.61)	6.48 (1.01)
	8	69 / 68	7.26 (0.80)	5.91 (0.74)	6.57 (0.65)	6.38 (0.55)	5.93 (1.18)

Note. Psy_{Dutch} = Dutch psychology students; Psy_{Int} = International psychology students; ECS = Education and Child Studies students; $N_{evaluation}$ = Number of students who filled in the course evaluation; N_{grade} = Number of students who obtained a grade for the course knowledge assessment; GG_{aim} = aimed grade goal; $GG_{minimum}$ = minimum grade goal; PSE_{course} = course performance self-efficacy; PSE_{year} = year performance self-efficacy.

Table 2. Pearson correlations between the motivational variables, as well as with course grade, per course.

	$GG_{minimum}$	PSE_{course}	PSE_{year}	Grade
GG_{aim}	*1.* .39	1. .52	1. .33	1. $.05^{ns}$
	2. .38	2. .50	2. .34	2. .26
	3. .64	*3.* .71	3. .39	3. .37
	4. .50	*4.* .59	4. .38	4. .19
	5. .55	*5.* .58	5. .42	5. $.08^{ns}$
	6. .52	*6.* .62	6. .34	6. .17
	7. .48	*7.* .58	7. .35	7. .22
	8. .48	*8.* .54	8. .31	8. .21
$GG_{minimum}$		1. .53	*1.* .40	1. $.00^{ns}$
		2. .46	2. .37	2. .19
		3. .69	3. .32	3. .32
		4. .52	4. .39	4. .22
		5. .63	5. .51	5. $.14^{.01}$
		6. .53	6. .34	6. .24
		7. .61	7. .38	7. .21
		8. .57	8. .38	8. .21
PSE_{course}			*1.* .49	1. $.03^{ns}$
			2. .46	2. .28
			3. .42	3. .43
			4. .36	4. .23
			5. .56	5. .17
			6. .45	6. .19
			7. .50	7. .26
			8. .46	8. .24
PSE_{year}				1. $-.01^{ns}$
				2. .19
				3. .26
				4. .35
				5. .25
				6. .40
				7. .43
				8. .40

Note. All correlations were significant at level $p < .001$, except the correlations flagged as .01 ($p < .01$) , or n.s. (non-significant: $p > .05$); GG_{aim} = aimed grade goal; $GG_{minimum}$ = minimum grade goal; PSE_{course} = course performance self-efficacy; PSE_{year} = year performance self-efficacy.

Latent Classes of Motivation (RQ1)

Number of Classes

In Table 3 we present the fit indices, entropy and average class probabilities of the different models. Results for the mixture modelling showed that a 2-class model was a significant improvement over a 1-class model. Thereafter, the addition of extra classes did not significantly improve model fit: the VLMR and LMR were non-significant for the 3- and 4-class models. The remaining criteria confirmed the superior fit of the 2-class model. First, the percentages of students per group for the 2-class model were substantial: 65.6% and 34.4%, respectively. Secondly, the 2-class model had the highest entropy (.90), which indicates that within-class differences were relatively small compared to between-class differences. Thirdly, the average class probabilities were close to 1.00, indicating that students could be classified with high confidence. In conclusion, a 2-class model was the best fit to the data.

Results from the multi-group analyses showed that the two-class model was invariant for the three subsamples (Psy-Dutch/Psy-International/ECS). The comparison of the model with fixed items means versus freely estimated item means yielded a significant difference, Satorra-Bentler χ^2 (130) = -544.58, $p < .001$. However, the fact that this χ^2 was negative indicates that the model with fixed items means has better fit. The comparison of the model with fixed class probabilities versus freely estimated class probabilities yielded a non-significant Satorra-Bentler $\chi^2(2) = 0.29$, $p = .863$. Therefore, we did not need to differentiate between the three subsamples in our analyses.

Table 3. Results for the mixture modelling (RQ1): VLMR, LMR, entropy and average class probabilities per model.

Model (comparison)	ΔDf	VLMR (*p*-value)	LMR (*p*-value)	Entropy	Average class probabilities
2-class (vs. 1-class)	33	3498.01 (.023)	3481.46 (.023)	.90	.974/.963
3-class (vs. 2-class)	33	1141.50 (.092)	1136.10 (.093)	.89	.954/.956/.934
4-class (vs. 3-class)	33	420.18 (.501)	418.19 (.503)	.88	.949/.891/.929/.961

Note. Average class probabilities describe the average probability of belonging to a certain class, for all students being assigned to that class. Thus, the number of probabilities given corresponds to the number of classes in the model.

Class Interpretation

For a visual illustration of the two classes, see Figure 1. Class 1 contained 65.6% of the students. Although aimed grade goals were higher than the 6.0 performance standard, class 1 students on average indicated that their minimum grade goals were near the 6.0 performance standard. In addition, class 1 course performance self-efficacy and year performance self-efficacy were only slightly above this standard as well. As for the entire group of students, the only exception was course 3 (statistics): here, students showed a drop in grade goals and course performance self-efficacy. In sum, overall class 1 students' grade goals and performance self-efficacy were close to the performance standard. Consequently, we characterised class 1 as the *sufficient motivation* class.

Class 2 contained 34.4% of the students. Class 2 students held aimed grade goals that were about 1 point higher on average than the *sufficient motivation* class. For class 2, the minimum grade goal was higher than the 6.0 performance standard in all courses except course 3. Class 2 had higher course performance self-efficacy and year performance self-efficacy than class 1, which indicates that class 2 expected higher grades. As for class 1, grade goals and course performance self-efficacy dropped in course 3. Overall, as can be seen in Figure 1, the two classes showed a similar pattern over the eight courses. However, class 2 had higher average grade goals and self-efficacy on all eight courses. Therefore, we described class 2 as the *high motivation* class.

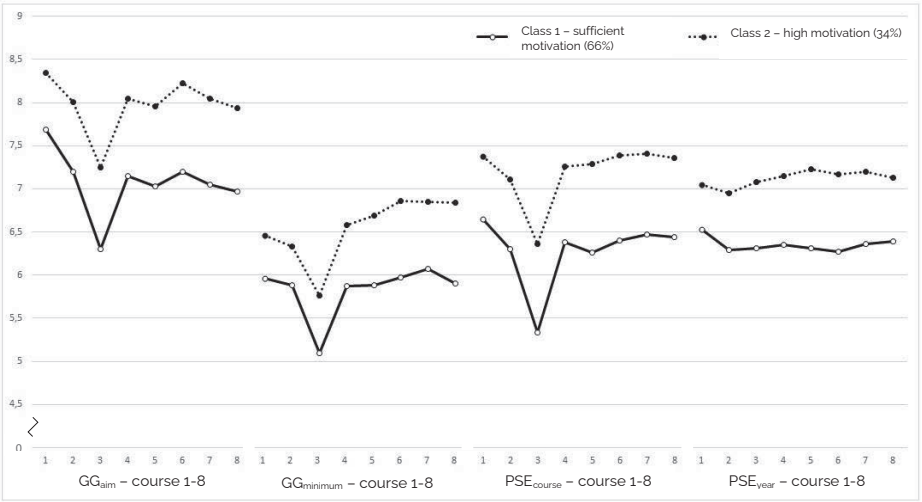


Figure 1. Average item scores for the four motivational variables on the eight first-year courses for the two latent classes. GG_{aim} = aimed grade goal; $GG_{minimum}$ = minimum grade goal; PSE_{course} = course performance self-efficacy; PSE_{year} = year performance self-efficacy.

Table 4. Descriptive statistics and significance for the dependent variables in the *t*-tests (RQ2)

Variable	Class 1 (*N* = 385) 'sufficient motivation'		Class 2 (*N* = 202) 'high motivation'				
	Mean	SD	Mean	SD	*t*	*p*	*d*
Use of time	13.38	6.37	14.61	7.61	1.96	.051	0.18
Interest	3.64	0.51	3.78	0.52	3.16	.002	0.27
Relevance	4.02	0.46	4.14	0.45	2.96	.003	0.26
Course rating	6.83	0.72	7.10	0.68	4.39	<.001	0.38
Amount of learning	3.97	0.45	3.97	0.43	0.00	.998	0.00
Average grade	6.22	1.00	7.05	1.01	9.51	<.001	0.83

Note. 1 student from class 1 had no grades, hence N = 384 for class 1 in the average grade t-test. SD = standard deviation; *t* = *t*-value; *p* = *p*-value; *d* = effect size (Cohen's *d*).

Further Class Characterisation (RQ2)

We performed six t-tests with class membership as the independent variable, and use of time, interest, relevance, course rating, amount of learning and average grades as the dependent variables. The *high motivation* class had a significantly higher mean than the *sufficient motivation* class on interest, relevance, course rating, and grades ($p < .01$), but the classes did not differ significantly in terms of use of time and amount of learning (see Table 4 for the results).

Discussion

The purpose of the current study was twofold. Our first aim (RQ1) was to identify latent motivational classes based on students' grade goals and performance self-efficacy. A two-class solution provided the best fit to the data. We differentiated between a *sufficient motivation* class that contained 66% of students, and a *high motivation* class with 34% of students. The second aim (RQ2) was to further delineate and validate the classes, by investigating whether the classes differed on a number of extraneous variables, which was indeed the case: The *high motivation* class scored significantly higher on several course evaluation items, as well as in terms of grades.

Two Latent Classes

In our analyses, we found evidence for the existence of two latent classes: a *sufficient motivation* class, and a *high motivation* class. Both classes of students held grade goals and performance self-efficacy at or above the performance standard. This is in line

with previous research (Kickert et al., 2019), and signifies that all students wish to pass assessments, and progress academically. The difference between the classes was that the *sufficient motivation* class had grade goals and performance self-efficacy that were near the performance standard, whereas the *high motivation* class had higher goals and self-efficacy.

The distribution of students over the two classes is informative: 66% were in the *sufficient motivation* class, and 34% were in the *high motivation* class. Thus, 34% of students did not seem to let the faculty determine their grade goal and performance self-efficacy, as these students showed a motivational profile higher than needed to pass the assessments. However, the grade goals and performance self-efficacy of the *sufficient motivation* class were close to the GPA performance standard, which indicates that 66% of students have grade goals and levels of performance self-efficacy that mirror the performance standards. In other words, the level of performance that these students were aiming for closely resembles the minimum level of performance needed to pass. Although this minimum passing level may be the best possible performance for a proportion of the *sufficient motivation* class, given the large number of students in this class, it seems unlikely that all these students were motivated to show their best possible performance. Thus, perhaps the current performance standard was not a big enough challenge for a proportion of the *sufficient motivation* class.

The distribution of students over the two classes adds to previous research that showed a relationship between characteristics of assessment policies, motivation and performance (Kickert et al., 2018, 2019). For instance, our results add nuance to the previous finding that students' academic performance seems highly responsive to the performance standard (Kickert et al., 2018). In the current study, we found that 66% of students held grade goals, performance self-efficacy and average grades near the performance standard. Thus, based on these results, we expect that the performance of up to 66% of students may be particularly responsive to changes to the performance standards.

Our analyses also indicated that the two-class solution was invariant for the three subsamples (i.e. Dutch psychology students, international psychology students, and ECS students). Although these subsamples consist solely of social sciences students, the invariance indicates that the two-class solution applies in different course programmes, and with different student populations. In sum, over the three subsamples, we found a robust two-class solution that consists of a *sufficient motivation* class and a *high motivation* class.

Motivational Patterns

The comparability of the motivational patterns of the two classes is striking, as the patterns of grade goals and performance self-efficacy throughout the year are almost exactly parallel. Students in both classes show a slight drop in grade goals and self-efficacy after the first course, and thereafter remain relatively stable over all courses except the third course, statistics. Thus, the major difference we observed between the two classes concerned the level of motivation, rather than class differences in developmental trajectories in terms of motivation.

The drop after the first course is congruent with the literature on the development of grade goals and performance self-efficacy within a course: students begin with overly optimistic grade goals and performance self-efficacy, and subsequently lower their goals and performance self-efficacy (Guillaume & Khachikian, 2011; Hossain & Tsigaris, 2015; Khachikian et al., 2011; Radosevich et al., 2004). The drop after the start of the year is also mentioned in literature on student confidence (Putwain & Sander, 2016). With respect to this drop, it is relevant to note that course performance self-efficacy was significantly correlated with course grades in all courses except the first. A possible explanation is that students needed the first assessment in order to form more accurate expectations of the assessments. These findings are in line with research showing that pre-university self-efficacy is not predictive of academic success (Van Herpen et al., 2017). Together, the findings concerning the first course seem to indicate that students have difficulties with estimating their capabilities at the start of their academic career.

Given the comparable motivational patterns, in both classes we observed evidence of context-specificity as well as stability of grade goals and course performance self-efficacy. Concerning context-specificity, in the statistics course (course 3) we observed a sharp drop in grade goals and course performance self-efficacy. The fact that this drop was paralleled by a drop in course grades reflects the accuracy of students' goals and expectations. We interpret the drop in motivation as an indication of the context-specificity of students' goals and expectations of performance.

However, in contrast to this context-specificity, students within both classes showed relatively stable goals and self-efficacy in seven out of eight courses. This stability may denote that students do not adapt their grade goals and self-efficacy to the courses and obtained grades, or that their grades actually match their goals and expectations. The observed correlations of the motivational constructs with grades, the stability of the year performance self-efficacy, and the drop in the statistics course point towards the latter explanation.

For many students, statistics courses are known to be associated with more test anxiety than other courses (Onwuegbuzie & Wilson, 2003). Thus, in the current investigation students only seemed to adapt their otherwise rather stable grade goals and performance self-efficacy in the 'extreme case' of a statistics course. Future research needs to show which strategies students adopt in determining their motivation, and how educational contexts affect these strategies.

Further Characterisation of the Classes

We obtained additional useful information for the characterisation of the classes, by relating the classes to several other variables. Additionally, the relationships with these other variables provide empirical evidence for the validity of the distinction between the two classes. Compared to the *sufficient motivation* class, students in the *high motivation* class did not spend more time on individual study. This contrasts with the literature that shows that students with high performance self-efficacy spend more time studying (Guillaume & Khachikian, 2011). For both classes, the average amount of time spent on self-study is lower than the twenty hours prescribed in both course programmes. In our view, these low averages underline the hypothesis that the current performance standard was not an abundant challenge for many students.

We did observe significant differences between the two classes on other variables: the *high motivation* class found the courses more interesting and relevant, and gave higher overall course ratings. Interest and relevance are considered to be components of the motivational construct task value (Pintrich et al., 1993). Therefore, despite the fact that the used student evaluation items are not validated measures of task value, we believe that the *high motivation* class experienced higher task value. Thus, although extrinsic rewards such as grades undermine intrinsic motivation (Deci et al., 1999), aiming for and expecting high grades goes hand in hand with valuing learning more.

There were no significant differences between the classes in terms of perceived amount of learning. A possible explanation is that students with higher grade goals and performance self-efficacy already enter the course with a higher level of knowledge. Alternatively, it may be that students experience difficulties in judging their amount of learning. Research shows that low performers particularly suffer from overconfidence about their competence (Kruger & Dunning, 1999). Thus, perhaps the *sufficient motivation* class overestimated their amount of learning.

Finally, the *high motivation* class showed significantly higher academic performance. These higher grades are in accordance with higher grade goals and higher performance self-efficacy. These associations are also reflected in the correlations we

observed between the grade goal items, performance self-efficacy items, and grades. The large effect size of the performance difference between the classes underlines the practical significance of the differentiation between the classes. In sum, the picture emerges of a *high motivation* group with relatively high goals and self-efficacy, who value their learning more and perform better than the *sufficient motivation* group.

Limitations

This research had several limitations that we want to address. Firstly, the motivational data, as well as the outcome variables used to characterise and validate the classes, came from student evaluations. On the one hand, this can result in socially desirable answers, despite the fact that students were informed that all data would be treated anonymously. On the other hand, some students may not have been inclined to invest effort in adequately filling in the evaluations.

Importantly, these student evaluation data were observational, hence no causal conclusions can be drawn on the basis of the current investigation. Additionally, the outcome variables used to further characterise the classes did not originate from validated questionnaires. Therefore, we have no available information on the validity and reliability of these items. Although the same principally holds for the grade goal and performance self-efficacy items, these (or highly comparable) items have already been used for many years (Locke & Bryan, 1968).

Finally, several aspects of the specific educational context may be relevant as well. Firstly, all data from the current study came from social science students. Secondly, these students all studied in a problem-based learning environment, characterised by small-group tutorial meetings and with only one knowledge course and one skills course in each block period. Thirdly, all assessments were high stakes. Perhaps different results would have been obtained in different disciplines, other types of learning environments, or in case of lower stakes assessments. The stakes may be particularly relevant, as the stakes are known to affect the association between goals and performance (Locke & Latham, 2002).

Future Directions

A first possibility for future research is to investigate whether similar results are obtained in other student populations, or in different curricula. For instance, different results may be obtained with medical students, in more traditional teacher-centred curricula, or under assessment policies with lower stakes. These different results could mean that different classes are identified, or that the distribution of students over the current classes differs.

A second possibility is to longitudinally investigate the classes' motivation and academic performance. For instance, do the students switch to another class in subsequent academic years, and why? And how do the classes perform academically in the following years?

A third possibility is to look for motivational variables that would further our understanding of the classes' motivation. For instance, adding more variables may result in subdivisions within the current two classes. As our motivational variables were all grade-related, we would be particularly interested to find out what proportion of students within both classes aims to learn unassessed objectives out of their personal interest. Perhaps some students in the *sufficient motivation class* aim to minimally pass in order to focus on their own interests, whereas other members of this class simply aim to minimally pass.

Conclusion and Implications

The most important conclusion of this research is that 66% of students held grade goals and performance self-efficacy beliefs that were very close to the performance standard of the assessment policy, and remained close to that standard over the course of an academic year. In other words, we suspect that these students were motivated to achieve the minimum grades needed to progress academically. Consequently, establishing a performance standard is a crucial motivational element of an assessment policy, or more broadly speaking, of a curriculum. Our results raise the question of what would happen when the performance standard would be higher, or perhaps more provocatively, when the standard would be unknown to the students. A first approach then would be to set the standard, without communicating it to the students. An alternative is not to set a standard beforehand, and let an expert panel make decisions on academic progress (Van der Vleuten et al., 2012). Ideally, students would aim for their maximum performance. As our results indicate that many students instead focus on the minimally sufficient performance needed to pass, teachers should carefully determine the best performance standards.

CHAPTER 6

Curricular fit perspective on motivation in higher education

This chapter is submitted as:

Kickert, R., Meeuwisse, M., Stegers-Jager, K.M., Prinzie, P., & Arends, L.R. (submitted). Curricular fit perspective on motivation in higher education

Abstract

In this paper we present a theoretical perspective that can be used to understand students' focus on assessment in higher education. We propose that the degree of alignment between the objectives and assessment of the curriculum plays a crucial role in students' motivation. In case of perfect alignment, all objectives have an equitable probability of being assessed. Thus, all learning contributes to performance equitably. Consequently, the motivation to perform and the motivation to learn should result in the same learning behaviour and performance. However, in reality a certain degree of cognitive and operant misalignment of the assessment with the objectives is present. Hence, some objectives will not need to be mastered in order to pass certain assessments. Consequently, a distinction arises between assessed and unassessed learning, and only the assessed learning contributes to performance. Thus, the probability of performing well on assessments is higher when students focus their effort on the assessed learning only, instead of dividing their effort between the assessed and unassessed learning. Therefore, students who are motivated to perform, have a motivation that fits in a misaligned curriculum. The paper concludes with implications of this curricular fit perspective for assessment practices, as well as for motivational research.

Keywords: motivation; higher education; alignment; curricular fit; motivation to learn; motivation to perform

Introduction

Most motivational researchers will agree that students learn best when learning tasks are considered enjoyable or interesting, as students who consider learning enjoyable or interesting will have mastery goals (Ames, 1992) and/or will be autonomously motivated (Ryan & Deci, 2000). Thereby, current motivational theories such as goal orientation theory (Ames, 1992; Dweck, 1986; Nicholls, 1984) and self-determination theory (Ryan & Deci, 2000) prescribe the optimal learning situation: educators should strive to make their curricula enjoyable or interesting in order to optimally motivate students.

However, most higher education students are more focused on their assessment performance than on their learning enjoyment or interest (Becker et al., 1968; Cilliers et al., 2010). This focus on assessment performance can take different forms: all students have the goal to pass assessments; a subsample of students aims for higher grades (Kickert et al., 2019). We suggest that the explanation for students' focus on assessment should not be sought in individual students, but in the description of which learning is rewarded by the curriculum, through grades. More specifically, we posit that students with a focus on assessment performance have a motivation that fits in a misaligned curriculum. The benefit of our proposed curricular fit perspective is that the focus shifts towards the way curricula shape student motivation, thereby leading to concrete suggestions for how to improve student motivation through adaptation of the curriculum.

Firstly, we describe how misalignment of assessment with the objectives of the curriculum comes to occur. Secondly, we explain how students come to expect this misalignment. Thirdly, we substantiate how expectations of misalignment affect student motivation, by discouraging students from learning unassessed content. Fourthly, we present several curricular adaptations that can help to motivate students to learn the full curriculum, instead of only the assessed curriculum. Finally, we present implications of this curricular fit perspective for motivational research.

Curricula and Alignment

Curricula consist of three primary elements: objectives, instruction (including both instructional activities and materials), and assessments (Anderson, 2002). The objectives determine the intended outcomes of learning, instruction is the means through which these objectives should be achieved, and assessments serve to determine whether the objectives are achieved. These three elements of the curriculum are the educators' means to affect student motivation.

Importantly, in order for education to be effective there should be *alignment* between objectives, instruction, and assessment of the curriculum (Anderson, 2002; Biggs, 1996; Cohen, 1987). Alignment means that there is congruence between the objectives, instruction and assessments. In laymen's terms: instruct what you intend to teach, and test what you have taught. This means that the assessment is a random sample out of the population of objectives (see Figure 1a). As a result, in case of alignment, all objectives have an equitable probability of being assessed. Hence, all learning behaviour that was intended by the curriculum has an equitable probability to aid students' assessment performance.

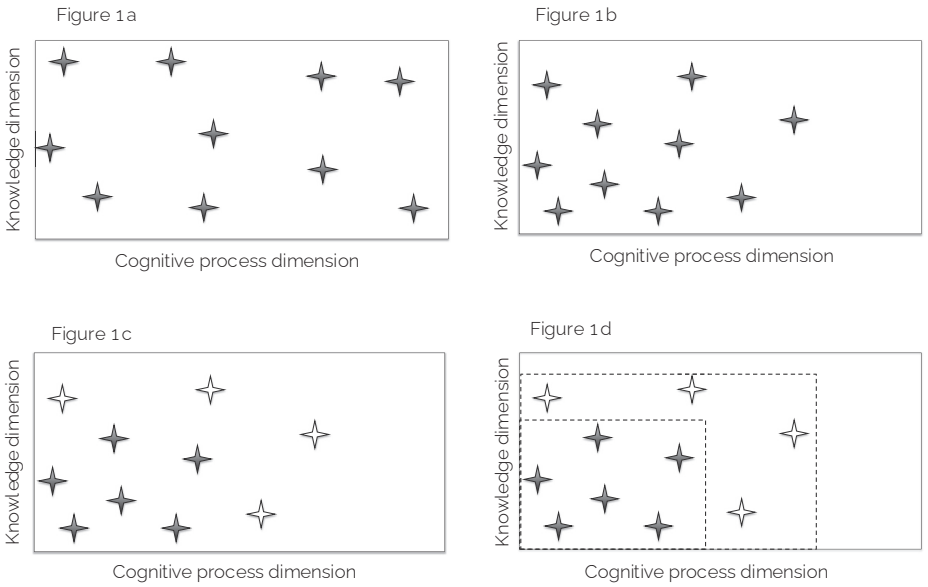


Figure 1. A schematic representation of the relationship between alignment and the distribution of effort for students who are motivated to perform. The square represents the curricular objectives, the stars represent the assessment items.
Figure 1a: alignment of objectives and assessment, as the items are evenly spread around the area;
Figure 1b: cognitive misalignment, as the stars are not evenly spread around the objectives square;
Figure 1c: cognitive and operant misalignment, answering the black stars (i.e. assessment items) correct is sufficient to pass the test.
Figure 1d: A Student who manages to focus efforts on the area within the dotted square will have better chances of performing well on the assessment items than when that student spreads efforts over the whole square. The larger dotted square represents a student who wants to get a perfect score on the assessment, the smaller dotted square represents a student who wants to pass with a sufficient grade.

However, in reality there will be a certain degree of misalignment of the assessment with the objectives. We will now explain why misalignment occurs, using the distinction that Cohen-Schotanus (1999) has made between cognitive and operant aspects of learning that are affected by assessment. Cognitive aspects concern the content of learning (i.e. what and how), and thus include the knowledge covered, as well as the required level of processing; operant aspects of learning refer to the amount of required learning (i.e. when and how much). We suggest that this distinction between cognitive and operant aspects can be extended to misalignment as well.

In case of *cognitive misalignment*, some objectives' content will be relatively underrepresented in the assessment. Krathwohl (2002) describes Bloom's revised taxonomy, in which the content of educational objectives can be represented in a knowledge dimension and a cognitive process dimension. The combination of these two dimensions results in an educational objective, wherein the knowledge dimension embodies the noun and the cognitive process embodies the verb. For instance, an objective for a social sciences curriculum can be that graduates can 'apply advanced statistical designs and methods', wherein 'advanced statistical designs and methods' is the knowledge, and 'apply' is the cognitive process. Cognitive misalignment occurs when certain knowledge or cognitive processing aspects of the objectives are inequitably represented in the assessment.

We identified several sources of cognitive misalignment based on the literature. As the whole is often greater than the sum of its parts, a first source of cognitive misalignment lies in the fragmentation of learning into smaller assessable elements (Lindquist, 1951; Sadler, 2007). Firstly, this fragmentation occurs because the curriculum is being divided into separate subjects, and assessment normally takes place at the subject level. Consequently, assessments concern the subject objectives, but not the curricular objectives. Therefore, ultimate learning objectives of the curriculum remain unassessed: "...the recognized ultimate objectives of instruction of individual subjects do not collectively constitute or account for the recognized ultimate objectives of the whole program of general education" (Lindquist, 1951, p.135). Secondly, within each subject the fragmentation continues, by deconstructing the subject objectives into smaller assessable elements, thereby further losing track of the greater whole (Sadler, 2007).

Besides fragmentation of learning into smaller assessable elements, other sources of cognitive misalignment are that some knowledge and skills are more likely to be assessed (Biggs, 1996; UNESCO, 2016), and that deep learning is often harder to assess than superficial learning (Frederiksen, 1984; Krathwohl, 2002). For instance,

objectives often concern integration and forming a substantiated opinion about the subject matter. However, in many cases, multiple-choice assessments are used for efficiency considerations, for example in case large groups of students need to be assessed. These multiple-choice assessments cannot assess whether the student can make innovative integrative connections or form an own substantiated opinion. Consequently, compared with an aligned curriculum, some aspects of learning will have an inequitable probability of being assessed in a cognitively misaligned curriculum. Because of this bias, the assessment will not be a random sample of the curricular objectives (see Figure 1b).

In case of *operant misalignment*, the amount of required learning for the objectives is larger than the amount of required learning for the assessment. Although the objective is for students to fully master a certain topic, a passing grade does not require fully mastering the topic. For instance, a passing grade often requires 50% to 60% correct answers on the assessment. Consequently, on the assessment, students can afford not to have mastered certain aspects of learning, and still obtain a passing grade (see Figure 1c).

In sum, due to cognitive and operant misalignment, some learning that is intended by the curriculum will not need to be mastered in order to pass the assessments. Thus, within misaligned curricula a distinction[1] arises between *assessed objectives* and *unassessed objectives*. Before we elucidate how this distinction may affect students' motivation, we will first discuss students' expectations of misalignment.

Students' Expectations of Misalignment

A necessary condition for misalignment to affect student motivation is that the student has *expectations of misalignment*. If a curriculum is misaligned, but the student does not expect misalignment, student motivation will not be affected. However, many students are aware that there is a conflict between learning and meeting the assessment demands (Becker et al., 1968; Cilliers et al., 2010; Öhrstedt & Scheja, 2018). In a seminal study, Snyder (1971) observed that students differentiate between the formal curriculum, and what he termed the hidden curriculum. The former contains the formal requirements of the curriculum, whereas the latter denotes what is actually expected in order to perform academically (Snyder, 1971). The crucial element in the hidden curriculum is assessment (Sambell & McDowell, 1998). In addition, improving alignment is associated with improved satisfaction among students, and with an increase of the

1 This is not a binary dichotomy, but rather a continuum of objectives with a very high probability of being assessed on the one end, and objectives with a very low probability of being assessed on the other end.

desired learning activities (Driessen & Van Der Vleuten, 2000; Newble & Jaeger, 1983). In sum, students seem to expect misalignment (Becker et al., 1968; Cilliers et al., 2010; Öhrstedt & Scheja, 2018; Snyder, 1971), and respond to it.

We can conceive of two sources of information that shape students' expectations of misalignment. A first source of expectations of misalignment can be students' previous experiences, both in preceding, misaligned subjects of the students' current curriculum, as well as earlier in a student's educational career (Boud, 1995; Sambell & McDowell, 1998). A recent study has shown that although students are not able to accurately predict their first grade at the university, predictive ability already improves considerably for the second grade (Kickert et al., 2020). Apparently, the first assessment helps to properly manage expectations of the assessments in the course programme. Additionally, previous experiences with assessments, also outside the curriculum, may have made the student aware that deep learning is difficult to assess. Therefore, the student can know that deep learning has an inequitable probability of being assessed.

A second source of expectations of misalignment are the implicit and explicit cues given about the assessment by the teacher. Research has shown that many students seek cues about what is more likely to feature in assessments (Becker et al., 1968; Cilliers et al., 2010; Miller & Parlett, 1974). Providing information on the assessments is often advised (Baartman et al., 2007; Broekkamp & Van Hout-Wolters, 2006), or even compulsory for teachers due to educational policy. Teachers may (be required to) communicate the assessment format during the subject, and knowing expected demands of assessments may affect students' learning (Baeten et al., 2010; Cilliers et al., 2010). For instance, students show differences in learning on multiple choice assessments versus essay assessments (Scouller, 1998; Stanger-Hall, 2012; Struyven et al., 2005), or on open-book versus closed-book assessments (Heijne-Penninga et al., 2008), and the type of assessment questions is associated with whether students aim for surface or deep learning while studying (Entwistle & Entwistle, 1991; Öhrstedt & Scheja, 2018; Struyven et al., 2005). For instance, when students know that the assessment will consist of questions that require reproduction of knowledge, students will aim for reproduction instead of transformation of knowledge, while studying (Entwistle & Entwistle, 1991). As a consequence, the amount of effort put into learning is related to the type of assessment; students invest more effort when the assessment is deemed relevant (Preston et al., 2020). Finley and Benjamin (2012) have even shown that students adapt their memory encoding strategy to the expected demands of an upcoming assessment, by experimentally demonstrating that students perform better when the assessment type was as expected, regardless of what that type was. As a

likely consequence of these adapted learning behaviours, students who expect to be assessed through assessments that require higher-order thinking skills, have a deeper understanding of the subject matter (Jensen et al., 2014), and test performance is best when students receive the kind of assessment they expect (Lundeberg & Fox, 1991; McDaniel et al., 1994; Thiede et al., 2011).

In addition to cues about test format, students will often be aware that less than 100% mastery is sufficient to pass an assessment, as the passing grade is another cue that is generally known prior to the assessment. Teachers can also provide practice exams, or make past exams public (Öhrstedt & Scheja, 2018). Reviewing past exams has been identified as an important cue seeking strategy that is associated with higher performance (Sebesta & Bray Speth, 2017). Additionally, material that is discussed in the lectures is deemed more likely to be assessed, especially in case of high frequency and intensity with which the material is discussed (Cilliers et al., 2010; Öhrstedt & Scheja, 2018). In summary, students have a host of informational sources to form expectations of misalignment.

We posit that the accuracy of expectations of assessment is a crucial determinant of academic performance, as students whose expectations are correct, have a strong advantage over students with misguided expectations: a correct expectation of what will be assessed can help in distributing effort towards the assessed learning, and will therefore positively impact performance. Now that we have described how misalignment occurs, and how students come to expect this misalignment, we will explicate what kind of student motivation fits in a misaligned curriculum.

Curricular Fit

Motivation to Learn and Motivation to Perform

Studying can serve many different ends for students, such as to get an interesting job, to get a high-paying job, to impress others or themselves, to become an expert, to feel the pleasure of learning, or to feel smart. However, within each individual subject, students have only two means to achieve these ends: through learning the subject materials, and/or through performing well on the subject assessment. Thus, we posit that within each subject, students can have two motivations for studying. The *motivation to learn* concerns the extent to which students aim to master curricular knowledge and skills. The *motivation to perform* concerns the extent to which students aim to perform on the assessment, i.e. the grades students aim for.

The motivation to learn and motivation to perform resemble self-determination theory's distinction between intrinsic and extrinsic motivation, respectively (Ryan & Deci, 2000).

Intrinsic motivation refers to performing an activity for the inherent satisfaction of the activity itself, whereas extrinsic motivation concerns performing an activity for some separable outcome. However, within self-determination theory, the focus is on the reasons students have to learn and perform, whereas we solely focus on the extent to which students want to learn and perform. In the following, we will use our distinction between the motivation to learn and the motivation to perform to elucidate how curricula shape students' motivation.

Motivation in a Misaligned Curriculum

We assume that all students aim to graduate. In order to graduate, students need to pass assessments. Hence, it has been observed that the lowest grade that students would be satisfied with, is never below the passing grade; regardless of what that passing grade is (Kickert et al., 2019). In other words, although studen©ts differ in which grade they are satisfied with, all students are motivated to perform.

In an aligned curriculum, all learning has an equitable probability to benefit performance on the assessments. As a result, in terms of performance, whether students are motivated to learn, and/or motivated to perform, will not matter as learning is a prerequisite to perform. In other words, in a perfectly aligned curriculum, all learning contributes to performance. Thus, the motivation to perform should essentially result in the same learning behaviour as the motivation to learn, and vice versa.

Conversely, in a misaligned curriculum, learning *assessed objectives* (we will refer to this as *assessed learning*) is profitable for assessment performance, but learning *unassessed objectives* (we will refer to this as *unassessed learning*) is not. Consequently, a student who is able to focus his or her effort on assessed learning will have better chances of performing well on assessments, compared to when that student would evenly spread his or her efforts among assessed and unassessed learning. As in higher education, grades are students' only formal and institutionalised reward for learning (Becker et al., 1968), students are only rewarded for putting effort in assessed learning.

Furthermore, given that students' time and energy are limited resources, focusing efforts on unassessed learning reduces efforts towards assessed learning, and thus should reduce assessment performance. Therefore, in terms of performance, a student is discouraged from putting efforts in unassessed learning because this lowers the chances of performing well on assessments. Indeed, Senko and Miles (2008) have reported that students who focus on personally interesting materials instead of on what the teachers find important, achieve lower grades than students who focus on what the teachers find important.

In addition to rewarding assessed learning and discouraging unassessed learning, in many higher education institutes there is another forceful mechanism at play: failing assessments can lead to academic dismissal (Stegers-Jager et al., 2011), which means that students with insufficient assessment performance are selected against. Thereby, in misaligned curricula, assessments are the motivational bottleneck: If a student does not pass the assessments, all other goals (e.g. learn how to become a good doctor/psychologist) are rendered useless as well. And indeed, students are aware that they need to survive in the short term by passing assessments, in order to reap the long term benefits of their education (Cilliers et al., 2010).

In sum, for students in a misaligned curriculum, regarding assessment performance it is maladaptive to distribute efforts towards unassessed learning, and adaptive to focus effort on assessed learning (see Figure 1d). We use the term *adaptive* to underscore the fact that students who are motivated to perform have a motivation that *fits* in a misaligned curriculum. This motivation should positively impact higher education students' only form of formal rewards: grades. In addition, good grades increase the students' chances of 'survival', i.e. passing assessments. Conversely, students who are not motivated to perform, are more likely to fail assessments and face academic dismissal, and thus are 'selected against'. Therefore, the larger the misalignment, the more adaptive it will be for students to be motivated to perform.

An Analogy: Training for a Marathon

As an analogy, suppose Sarah is motivated to perform well on a marathon. This is an example of a situation that should have excellent alignment, as the objective (i.e. run a marathon) is congruent with the assessment performance (i.e. finish time on a marathon). Now suppose Sarah knows the assessment will be misaligned; her marathon performance will only be assessed by measuring her time on the first half marathon. If Sarah wanted to perform as well as possible on this assessment, she would probably adapt her training to this shorter distance. The shorter this assessed distance becomes, the larger the misalignment, and the more this would affect her preparation. In an extreme case of misalignment, suppose her marathon performance would only be assessed over a hundred meter interval. Her training would likely feature an excessive amount of explosive sprinting, and Sarah would perform much better than when she really would have trained for a full marathon. Her expectations of the assessment would have changed her assessment preparation, and thus gave her an advantage on the assessment. This change in preparation is an adaptive response to the misaligned assessment.

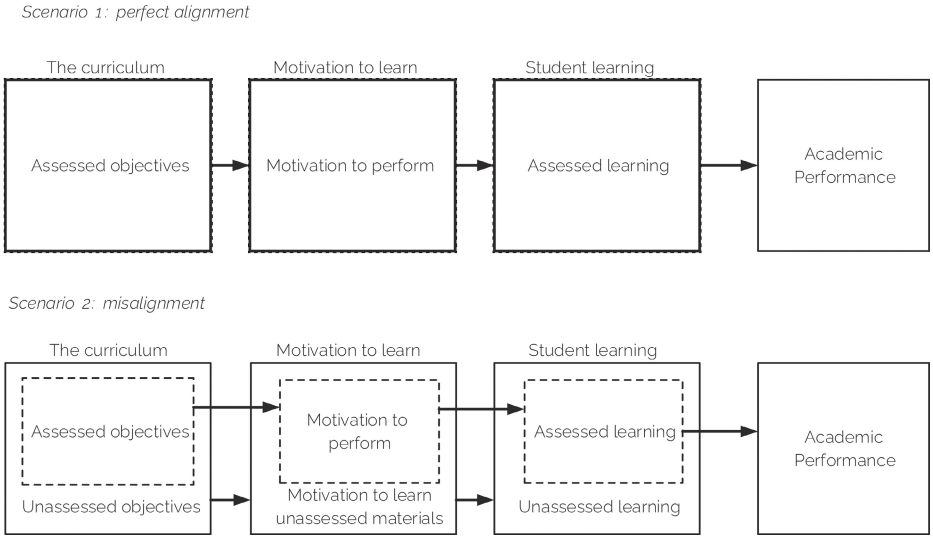


Figure 2. A conceptual model of student motivation in higher education. The degree of misalignment is represented by the surface of the unassessed objectives; in case of perfect alignment, this surface is non-existent (scenario 1). Hence, all objectives are assessed, and motivation to learn equals the motivation to perform. In case of misalignment (scenario 2), the unassessed learning does not contribute to academic performance. Consequently, the motivation to perform no longer equals the motivation to learn, and it becomes adaptive for a student to focus on the assessed learning only.

Figure 2 depicts a visual summary of our perspective of student motivation. In sum, the curriculum is the educator's tool to motivate students. As long as assessments are aligned with the objectives, it makes no difference whether students are motivated to learn or motivated to perform; both motivations will lead to the same performance. However, when misalignment occurs, in terms of performance, the motivation to perform becomes more adaptive than the motivation to learn. Hence, a misaligned curriculum is implicitly encouraging students to refrain from putting effort in unassessed learning.

Implications for Education

In summary, curricula consist of objectives, instruction and assessment. Through the curriculum, students can be motivated to learn, and motivated to perform. The cognitive and operant misalignment of assessment with objectives has important consequences for the most adaptive way for students to perform well. In case of alignment, the motivation to learn and motivation to perform have the same adaptive value for students who aim to graduate. However, the larger the misalignment, the more adaptive it becomes to be motivated to perform, instead of motivated to learn.

Students who are motivated to perform therefore have a motivation that fits better in a misaligned curriculum than students who are motivated to learn.

In a misaligned curriculum, assessment is the motivational bottleneck. By bottleneck we mean that motivating instruction will be of little use to motivate students in the long run if the assessments are not aligned with the objectives. Students may be motivated by the instruction initially, but once students find out that not all learning will be assessed, their motivation will narrow down as much as possible towards only the assessed learning.

Consequently, students' motivation is a reflection of the curriculum. For instance, if many students are not motivated to master all objectives, think critically, or show deep processing, the most likely explanation is that the curriculum is not motivating students the right way. Analogously, when scoring an exam, students' mistakes can be seen as a sign of what students need to learn better; however, if many students make a certain mistake, this should be seen as a sign of what the teacher needs to teach better.

Although we have explained that we believe all students are motivated to perform, we are not postulating that all students *only* want to pass. Many students wish to perform better than satisfactory (Kickert et al., 2019), and we believe that some students will want to put effort in unassessed learning, despite the curricular pressure to refrain from doing so. In fact, unassessed learning should be highly salient for students who are mindful of long-term benefits of learning. However, these long-term benefits can only be achieved in addition to the short-term goal to perform, because poor performance can lead to academic dismissal (Stegers-Jager et al., 2011). As assessments often serve to eliminate poor performers, many students are in a survival mode (Backer & Lewis, 2015). In other words, not all students may have the luxury to invest in long-term benefits, because these students are merely trying to survive.

We also want to explicitly state that we do not advocate an educational system that ignores students' interest, enjoyment or enthusiasm for learning. On the contrary: the tremendous benefits of enjoying an activity (Woolley & Fishbach, 2017), or of being intrinsically motivated (Cerasoli et al., 2014) are not under dispute. The point we have tried to make is that many contemporary curricula only reward learning assessed materials, and thereby implicitly discourage students to learn unassessed materials. Our assessment-minded educational system is pressuring students to be primarily motivated to perform.

When our curricula are indeed implicitly encouraging students not to invest effort in unassessed learning, the consequences for both students and society will be dire. Due to the focus on assessment, learning that is not (as easily) assessable runs the risk of not being done (UNESCO, 2016). As a consequence, students will graduate, but lack crucial knowledge and skills. We can conceive of a number of options to remedy this problem. A first route would run through the students; making students aware of the consequences that misalignment has for them could help students to focus on the long-term positive consequences of learning unassessed materials. However, as all students still need to pass assessments in the short term, increasing students' awareness of misalignment may also increase students' allocation of efforts towards the assessed learning, and thereby aggravate the adverse effects of misaligned curricula. Therefore, solutions need to be sought in the curriculum.

First Solution: Abandoning Assessment or Grades

A drastic option is to abandon assessment altogether (Becker et al., 1968). However, this would lead to an educational situation in which there is no standardised information available about the level of students' knowledge and skills. In addition, assessments can of course also be motivating for many students. Therefore, many educators will not find abandoning assessment a realistic option. However, as we have tried to substantiate above, a poorly aligned assessment can have adverse effects: although the assessment gives some information about students' knowledge and skill levels, it discourages students from performing the unassessed learning. Thereby, the learning process is corrupted. Hence, the adverse effects of misaligned assessments are not to be underestimated, and abandoning these assessments should be considered.

Instead of abandoning assessment altogether, we could reconsider the attachment of grades to students' performance (Tannock, 2017). Assessment is not equivalent to giving grades. In fact, Sadler (2014) asserted that codification of learning into the form of grading is impossible, even for pass/fail grading. Therefore, educators could give qualitative judgements, such as a verbal description of students' understanding of different topics, instead of grades. For an explanation of the reasons behind and method for qualitative judgements in workplace learning, see Govaerts and Van der Vleuten (2013).

Alternatively, we could strive to lower the importance of grades. According to Campbell's law, "the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor" (Campbell, 1976, p.49). Thus, reducing the consequences of grades could prove beneficial to our

educational system. A practical way to lower the stakes for individual assessments is to assess more often, with resulting lower stakes attached to each individual assessment (Van der Vleuten et al., 2012). This does not mean 'assess more', but 'assess smaller portions, more often'.

Second Solution: Improving Assessment

In addition to (partly) abandoning assessment or grades, assessment practices can be improved. First and foremost, this means we should strive to optimise alignment in our curricula. In essence, aligning assessments with curricular objectives means that the learning behaviour that was intended by the curricular objectives is rewarded by the assessments. A practical tool that can be used to assess cognitive alignment, is Bloom's revised taxonomy (Krathwohl, 2002). Both the educational objectives and the assessments can be placed in a table that consists of the knowledge dimension and cognitive process dimension of Bloom's taxonomy (Anderson, 2002). Then, the tables for the objectives and assessments can be compared in order to see which objectives are underrepresented in the assessment. Regarding operant alignment, educators need to assess whether the performance standards (i.e. grade required to pass) on the assessments are appropriate to determine whether the objectives have sufficiently been mastered. A necessary condition to improve alignment would be that educators receive the appropriate training, and are granted enough time to invest in improving their assessment practices.

Secondly, we should raise educators' awareness of the fact that assessments are a fundamental part of our curricula, and thus serve more purposes than measurement alone (Boud et al., 2018; Schuwirth & Van Der Vleuten, 2004). In particular, despite the strong traditional focus on assessment's reliability and validity (Boud, 1995), educators should also be aware of the motivational consequences of assessments. If the assessment solely rewards superficial learning, students are implicitly discouraged to perform deep learning. Gibbs and Simpson (2005) have even argued that "...we should design assessment, first, to support worthwhile learning, and worry about reliability later" (p.3). One way to support variation in learning, is to increase the variation in types of assessment (Broekkamp & Van Hout-Wolters, 2006). For instance, assessments of individual subjects which only concern subject objectives, could be supplemented by assessments of curricular objectives, such as progress testing (Van der Vleuten et al., 1996).

Thirdly, the prevailing view on assessment is one of damage control, in which assessments serve to exclude poor performers (Backer & Lewis, 2015). However, assessments also have the potential to inform, to make students push their boundaries,

and to be a force of positive change. In other words, educators need to reflect on whether they are assessing to find out what students do not know, or in order to elucidate what students do know. Again, this entails a shift from seeing assessments merely as evaluative tools, towards seeing assessments as educational tools.

Fourthly, instead of viewing assessment as the 'finish line' of a subject, the importance of the 'cooling down' could be reconsidered. Less cryptically, this means that exam reviews could be made a more fundamental part of the curriculum. Then, instead of students just knowing their grades, students could regularly reflect on which content was or was not mastered, based on the assessment performance. Which questions were answered correctly, which were not, and why? Making this reflection a customary part of the curriculum could aid all stakeholders in realizing that each assessment is not the endpoint of the learning experience, but a checkpoint somewhere along the way. Consequently, the distinction between formative and summative assessment, i.e. assessment for learning and assessment of learning, would cease to exist (Taras, 2005). In essence, all the above-mentioned ways to increase the quality of assessment require an increased self-reflection among educators on the possible influences of their assessments. This reflection requires time and energy.

Third Solution: Counter Strategic Effort

Given that perfect alignment often is an overly optimistic goal, a final resort may be to make it harder for students to be strategic in allocating their effort towards assessed learning, and not towards unassessed learning. Although transparency is often considered a quality criterium for assessments (Baartman et al., 2007), explicating detailed and transparent criteria of assessment can lead to assessment completely dominating the learning experience: assessment as learning (Torrance, 2007). As a possible remedy, students' expectations of cognitive misalignment can be obstructed, simply by not telling them how they will be assessed. If students know as little as possible about the assessment, preparation and assessment behaviour cannot be adjusted to the expectations either. Cilliers et al. (2010) observed that students were less likely to neglect certain learning tasks in case the assessors were perceived as less predictable.

The expectations of operant misalignment can be obstructed as well, by not communicating the performance standard before the assessment. If students know that 60% of the assessment items needs to be correct to pass an assessment, preparation for the assessment can be adapted to this standard. For instance, deep learning can be omitted because the superficial learning will suffice for a passing grade. Although it seems fair to give students all available information, not communicating the performance

standard may actually stimulate students to unleash their full potential, instead of unleashing their potential up until the point that the educator deems sufficient. In this scenario, we implicitly assume that a quantitative performance standard is known beforehand, but it may even be considered to let go of quantitative strategies to summarise assessment data, and use expert judgement instead (Van der Vleuten et al., 2012). In short, the fact that assessment criteria should be explicit and clear to the assessor, does not mean these criteria should also be communicated to the student.

So, in terms of the marathon example, a first option would be to just let students run the marathon, without measuring the finish time. The second option is to measure someone's marathon aptitude by assessing the full marathon. However, if for some reason only an interval can be assessed, a lot of adverse effects of this misalignment could be circumvented by the third option: not informing the runners about which interval will be assessed, or what time is considered to be sufficient.

Implications for Motivational Research

A first implication of our perspective for research on motivation is that we expect that the adverse effects of assessment on motivation are a consequence of misalignment. A well-known observation in motivational research is that extrinsic motivators such as assessments seem to have detrimental effects on students' intrinsic motivation (Deci et al., 1999; Harlen & Crick, 2003). We have presented a possible mechanism through which these effects can occur, and thus hypothesise that the reason for these detrimental effects lies in the misalignment of assessment with the objectives. Thus, if the assessment is perfectly aligned, we predict that assessment will not damage motivation. This prediction can be empirically investigated by measuring students' motivation under various degrees of expected misalignment.

A second implication concerns the measurement of motivation. The two concepts *motivation to learn* and *motivation to perform* are highly similar to the concepts intrinsic and extrinsic motivation in self-determination theory (Ryan & Deci, 2000) and mastery and performance goals in goal orientation theory (Ames, 1992; Dweck, 1986; Nicholls, 1984). However, within these theories, motivation is measured by asking for the reasons students have to learn and perform. For instance, in the Academic Motivation Scale, an example item for extrinsic motivation is "Why do you go to college? In order to obtain a more prestigious job later on" (Vallerand et al., 1992). An example item for a performance goal from the Patterns of Adaptive Learning Scales is "One of my goals is to look smart in comparison to the other students in my class" (Midgley et al., 2000). Although in different ways, both scales focus on the reasons for pursuing certain educational activities. Instead, we suggest that what essentially matters is not why students are motivated, but

how much effort students are willing to invest. We believe our stance is supported by the fact that the two motivational factors that show the strongest association with academic performance, are students' performance self-efficacy and grade goals (Richardson et al., 2012; Schneider & Preckel, 2017). Performance self-efficacy refers to the grades students expect to obtain, and grade goals are the grades students want to obtain (Richardson et al., 2012). These two factors both concern the 'how much' of motivation, instead of the 'why'.

The third implication for research also concerns the measurement of motivation. In a (hypothetical) perfectly aligned curriculum, the assessment is a perfect reflection of learning. Thus, in order to measure motivation, researchers only need to measure the motivation to perform *or* the motivation to learn, as both motivations will result in the same learning behaviour. However, in a (realistic) misaligned curriculum, researchers need to differentiate between the motivation to perform and the motivation to learn. The motivation to perform concerns the answer to the question 'to what extent do you want to do the assessed learning?'. The motivation to learn is essentially about answering the additional question 'to what extent do you want to do the unassessed learning?'; see Figure 2 for a visual illustration. However, asking this second question means we would assume that students are perfectly aware of misalignment. Therefore, an essential question that needs answering first is 'how well is the student able to predict which learning will be assessed and which will not?'.

Conclusion

In conclusion, we have presented a curricular fit perspective on motivation in higher education, by which we explain why it is more adaptive for students to be motivated to perform than to be motivated to learn in a misaligned curriculum. As we have attempted to substantiate, this perspective has implications for our assessment practices, as well as for motivational research. In our view, the most important benefit of conceptualizing motivation from a curricular perspective is that this puts the focus on those aspects of motivation that we can improve through our curricula. Educators have the privilege to shape these curricula, and thereby create their students' motivational context. Consequently, there are no good or bad kinds of motivation, just good or bad curricula.

CHAPTER 7

Summary and Discussion

The aim of this dissertation was to investigate whether academic progress and academic performance are associated with characteristics of assessment policies, and how motivation and self-regulation may explain possible differences in academic performance. Improving progress and performance is a continuous challenge for educators, and motivation and self-regulation are two of the most important factors associated with academic performance (Richardson et al., 2012). Yet, there is scarce literature available about progress, performance, motivation and self-regulation under different assessment policies. Therefore, the major assessment policy changes at Erasmus University Rotterdam (EUR) provided a unique opportunity to investigate the consequences that changes to the stakes, performance standards and resit standards have for student learning (i.e. progress, performance, motivation, self-regulation). In this dissertation, stakes are defined as the consequences of failing assessments, performance standards concern the passing grades needed to obtain credits, and resit standards concern the number of permitted resits. In this final chapter, we present a discussion of the major strengths and limitations, as well as educational implications and directions for future research. We first give a summary of the four empirical chapters and the theoretical chapter and indicate how our findings fit into the conceptual model (Figure 1) discussed in chapter 1 of this dissertation.

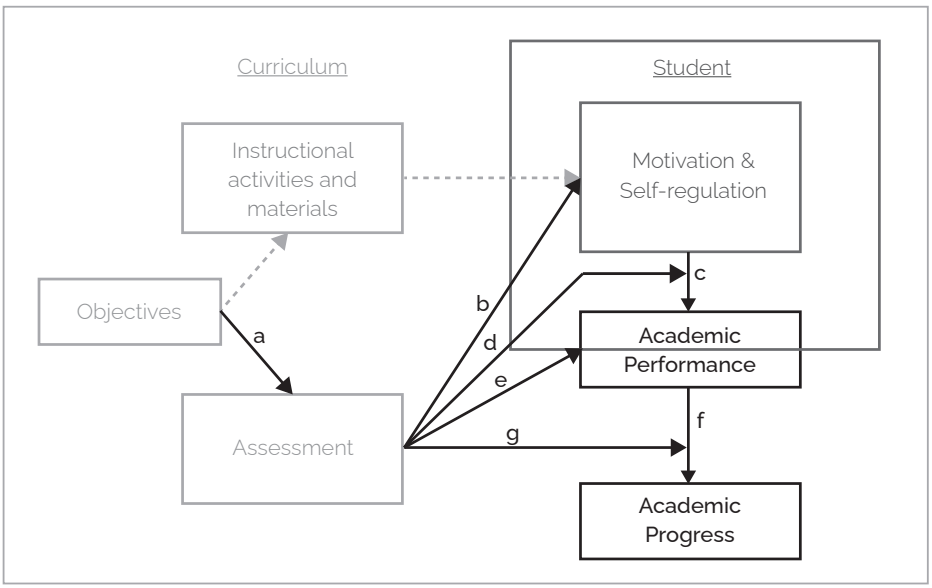


Figure 1. The general conceptual model for this dissertation. The black paths a-g represent the investigated associations in Chapters 2-6. Chapter 2 concerns paths e, f and g. Chapters 3 and 4 both concern paths b, c and d. Chapter 5 concerns paths b and c. Chapter 6 concerns paths a, b and c. The dotted paths are crucial for student learning, but were not the topic of this dissertation.

Summary

Explaining Progress: Differences in Performance and in Selection for Progress

The most important reason for changing the assessment policies was to accelerate students' academic progress. Therefore, the first research question (RQ1) of the study in **chapter 2** was 'What is the relationship between differences in assessment policies and differences in academic progress?' (*paths e, f and g* of Figure 1). Changing the assessment policy may however affect progress in two ways. Firstly, students under different assessment policies may show different academic performance, i.e. achieve different grades (*paths e and f*). For instance, when the stakes are raised, students may put in more effort, obtain higher grades, and consequently show higher progress. Therefore, RQ2 was 'What is the relationship between differences in assessment policies and differences in performance?'. Secondly, different assessment policies may result in a different selection for progress (*path g*). For instance, some students who would have progressed in the old policy may not meet the standards after the performance standard is changed. Thus, RQ3 was 'What is the relationship between differences in assessment policies and differences in selection for progress?'. To answer our three research questions, we investigated differences in first-year progress, performance, and selection for progress in three large course programmes at EUR: for business administration (n = 2,048) the main adaptation to the assessment policy was a change in stakes; medicine (n = 1,630) changed the stakes as well as the performance standard; psychology (n = 1,076) adapted the stakes, the performance standard and the resit standard. Although the changes to the performance standards and resit standards differed, the change in stakes was similar in all three programmes: in the old lower-stakes policies students needed to obtain all first-year credits within two years, whereas in the new higher-stakes policies these credits needed to be obtained within one year.

Therefore, regarding progress (RQ1) we compared progress after one year, as well as final progress, which was determined after two years under the old lower-stakes policy, and after one year under the new higher-stakes policy. Results showed that after one year, students in all three course programmes demonstrated significantly higher progress under the new higher stakes assessment policies. Thus, progress was faster under the new policy. However, final progress was lower under the new policy in business administration and medicine, yet higher in psychology. Hence, some students may not have been able to meet the higher demands of the new policy in two of the three course programmes. Particularly the significant differences in one-year progress differ from previous findings on academic dismissal (AD; Dutch: BSA) policies, which

found no association of AD policies with obtained credits or first-year completion rates (De Koning et al., 2014; Eijsvogels et al., 2015; Stegers-Jager et al., 2011). However, these previous studies compared low-stakes (i.e. no AD) with high-stakes policies (i.e. 2-year timeframe AD), whereas we compared high-stakes (i.e. 2-year timeframe AD) with even-higher-stakes policies (i.e. 1-year timeframe AD).

We offered two explanations for the progress differences. Firstly, these differences in progress may be explained by differences in performance (RQ2, *paths e and f*): students scored differently on several performance indicators under different policies. Our findings were generally in line with previous literature: higher stakes are associated with higher drop-out, but performance of the remaining students is higher (Lindo et al., 2010). However, conclusions vary depending on the investigated student population (i.e. all students or only progressors), educational context (i.e. course programme and type of assessment), as well as the type of performance indicator (i.e. average grades or mimicked progress). Secondly, differences in selection for progress may also account for progress differences (RQ3, *path g*): a significant proportion of students would progress under one policy but not under the other. Thus, it mattered for students which policy is used to decide about their progress.

In sum, results of this study indicate that both in terms of progress and performance, students seemed to be sensitive to the assessment policy. Thus, assessment policies may be an effective way to shape student progress, both because of differences in performance and differences in selection for progress.

Explaining Performance: Differences in Motivation and Self-Regulation

The next step was to explain the differences in academic performance between assessment policies. Therefore, in the following two chapters, we examined two of the most important predictors of performance: students' motivation and self-regulation (Richardson et al., 2012). In the study in **chapter 3**, we used a previously developed structural model of medical student performance that was developed and empirically tested under the old policy (Stegers-Jager et al., 2012). This model consisted of 'motivational beliefs' (i.e. motivation), 'learning strategies' (i.e. self-regulated learning; SRL), and 'participation in scheduled learning activities'. In this model, motivational beliefs had positive associations with performance, and this association was mediated by higher learning strategies and higher participation.

Our first research question (RQ1) was 'Are there differences in average scores of motivation, SRL, participation and academic performance of students under an old and a new assessment policy?' (*path b*). Therefore, we compared average scores on

the structural model's motivational and SRL constructs, as well as for participation and performance. RQ2 was 'Are the relationships between academic performance, motivation, SRL and participation similar across the old and new policy?' (*paths c and d*). To this end, we tested whether the structural model was invariant for students under both policies. We obtained official grades from university records and responses on a self-report questionnaire from students who studied either under a conjunctive lower stakes, lower performance standard (old) assessment policy (n = 648), or under a compensatory higher stakes, higher performance standard (new) assessment policy (n = 529).

Results for RQ1 showed that, although we did not observe significant differences on all observed variables, in general, motivation, SRL, and participation were significantly higher under the new higher stakes policy (*path b*). Regarding motivation, new policy students valued their learning more, and had more belief in their own competence. In terms of SRL, the new policy students generally showed more deep learning (e.g. relating material to what is already known and reflecting on their own learning) and better management of resources such as time and effort. Academic performance was also significantly higher under the new higher stakes policy. The performance difference between students under the two policies was strikingly similar to the difference between the performance standards of the old and new policy: half a point on a ten-point scale.

Concerning RQ2, the structural model was invariant between both policies, indicating that the relations between motivation, SRL, participation and performance were similar (*paths c and d*). Under both policies, higher performance could be explained by a positive path that starts from value, through deep learning, resource management, and finally participation. Our results are not in line with previous findings of Sundre and Kitsantas (2004) and Sungur (2007) who showed different associations of motivation and SRL with performance under different stakes. However, these previous investigations compared low-stakes with high-stakes assessments (i.e. no consequences vs. consequences), whereas we compared high-stakes to higher-stakes (i.e. consequences vs. even higher consequences). We concluded that students' higher performance under the new policy could be explained by higher motivation, SRL and participation, but not by different associations between motivation, SRL, participation and performance.

In the study in **chapter 4** we also investigated motivation, SRL and academic performance under two different assessment policies, yet now using social sciences students and the strongest motivational and SRL predictors of performance, as

identified in the meta-analysis by Richardson et al. (2012). RQ1 was 'Can we replicate the earlier reported finding that academic performance is higher under more difficult assessment policies?'. Therefore, we compared official grades from university records on the initial assessment, the number of times a resit was taken, as well as final official grades (after the resit), for two groups of students in the same single statistics course. We utilised the fact that this course was identical for second-year psychology students and third-year education and child studies (ECS) students, whereas the assessment policy differed for both groups of students: psychology students (n = 219) performed under an assessment policy with lower stakes, a lower performance standard and a higher resit standard than ECS students (n = 85). Therefore, we could compare students under two different assessment policies within the same single course. RQ2 was 'Are there average differences in the strongest motivational and SRL predictors of academic performance under both assessment policies?' (*path b*), and RQ3 was 'Are the associations of motivation and SRL with performance different under both assessment policies?' (*paths c and d*). Therefore, a subsample of the students ($n_{psychology}$ = 150/n_{ecs} = 51) filled in a questionnaire on motivation and SRL.

In terms of performance (RQ1), ECS and psychology students did not differ significantly on the initial exam. However, ECS students made more use of the resit, and showed higher final (post-resit) performance. Thus, we observed that ECS students postponed their higher performance until the resit. Additionally, ECS students generally reported higher motivation and SRL than psychology students (RQ2, *path b*), although not all differences were significant. Regarding motivation, ECS students had significantly higher goals, competence beliefs, and valued their learning more. In terms of SRL, ECS students showed better time management, but also reported more test anxiety. Regarding the associations of motivation and SRL with performance (RQ3, *paths c and d*), we did not observe significant differences. In addition, we observed that four motivational and SRL predictors of performance explained a significant proportion of variance in performance: aimed grade goals (i.e. '*Which grade are you aiming for*?'), performance self-efficacy (i.e. '*Which grade do you expect to earn*?'), academic self-efficacy (i.e. *belief in competence for learning and performance*), and effort regulation (i.e. *persisting effort in challenging or boring situations*).

Thus, in chapters 3 and 4, we observed several significant associations of assessment policies with motivation as well as SRL (*path b*) in both medical and social science students. Based on both chapters, we conclude that the higher final performance in the higher stakes assessment policy is associated with higher motivation and SRL (*path c*), but the relations of motivation and SRL with performance are similar between policies (*path d*). In other words, the way in which motivation and self-regulation affect

performance does not seem to differ between policies. Both studies fill important gaps in the literature, as there is a scarcity of research on motivation and SRL under different assessment policies. Thereby, we have gained a better understanding of the mechanisms through which assessment policies are associated with performance. Most importantly, our observations indicate that the assessment policy can be harnessed to improve performance, through higher motivation and SRL.

Summarizing our conclusions from chapters 2-4, we have shown that students' progress was sensitive to characteristics of the assessment policies in various course programmes: when the timeframe to obtain credits is shorter (i.e. higher stakes), students show faster progress. These progress differences can be explained by differences in performance (*paths e and f*), as well as by different selection for progress (*path g*). We also concluded that these differences in performance can be explained by differences in motivation and self-regulation (*paths b and c*), but not by different relations of motivation and SRL with performance under different policies (*path d*). However, we suspected that not all students' motivation may have been affected by the assessment policy in a similar fashion (i.e. *path b* may differ). More specifically, we expected some students to be more focused on meeting the minimum performance standard than others. Therefore, this potential motivational difference between classes of students was the focus of the next chapter.

Does the Assessment Policy Affect all Students Similarly? Latent Classes of Motivation

The first research question (RQ1) of the study in **chapter 5** was 'Which latent classes of students exist in terms of the development of grade goals and performance self-efficacy throughout the first academic year?'. We selected grade goals and performance self-efficacy to measure motivation, as these are the two motivational factors most strongly associated with academic performance in higher education (Richardson et al., 2012; Schneider & Preckel, 2017). To our knowledge, this study was the first investigation of the development of these motivational constructs throughout an entire academic year. As the first year is crucial in higher education due to the academic dismissal (AD) policy, students' motivation during this year is highly important. We performed this investigation using a latent class approach, as this approach has the advantage that unobserved differences between students can be taken into account by forming unobservable (i.e. latent) classes of students based on observed variables. With the aim to further characterise and validate the classes, our second research question (RQ2) was 'How does latent class membership, based on grade goals and performance self-efficacy, relate with students' course evaluations and academic performance?'. To answer both RQs, we obtained social science students' (n = 587) self-reported grade

goals, performance self-efficacy, use of time, course interest, course relevance, course rating, and perceived amount of learning from students' course evaluations of all eight consecutive first-year courses (i.e. subjects). We obtained students' official grades from university records as well.

A two-class model showed superior fit to the data (RQ1). We termed class one (66% of students) the *sufficient motivation class*, as these students' grade goals and performance self-efficacy were close to the performance standard of the assessment policy throughout the entire year. Class two students (34% of students) had higher grade goals and performance self-efficacy than class one students in all eight first-year courses. Thus, we named this class the *high motivation class*. Both classes showed relatively stable grade goals and performance self-efficacy throughout the year, except for a drop in the single first-year statistics course. Thus, we observed both stability of motivation throughout almost an entire academic year, yet context-specificity of student motivation in the only substantively deviant course. The difference between the two classes did not concern the developmental trajectories of grade goals and performance self-efficacy, yet rather a difference in the level of motivation throughout the year. The validity of the two-class solution (RQ2) was supported by the fact that the *high motivation class* also reported higher course interest, course relevance, course rating, and achieved higher average grades than the *sufficient motivation class*. However, the classes did not differ significantly in terms of use of time and perceived amount of learning.

The main conclusion was that we suspected that students in the *sufficient motivation class* were motivated to obtain the minimum passing grades determined by the performance standard. Therefore, the study in chapter 5 adds to the previous studies in which we observed that students' learning appeared sensitive to the assessment policy; the current study suggests that the performance standard of the assessment policy was especially salient for 66% of the students. When the performance standard would be raised, these students would have to adapt their motivation, whereas the *high motivation class* students would not need to do so. Although the question remains what this percentage would be in other student populations, based on the observations in chapter 5 we hypothesise that the significant differences in progress, performance, motivation and SRL found in chapters 2-4 may chiefly have been caused by the *sufficient motivation class*.

In the four empirical studies described thus far, we observed that students in assessment policies with higher stakes and higher performance standards are generally more motivated, show better self-regulated learning, achieve higher grades,

and progress faster. Furthermore, we concluded that the assessment policy may be particularly salient for the *sufficient motivation class*. Although we are not surprised by our own observation that student learning seems sensitive to the assessment policy, we were surprised by the extent of this sensitivity. For instance, for the psychology students in the study in chapter 2, student progress in the new policy was as high after one year, as after two years in the old policy. And for medical students in chapter 3 we observed that the rise in performance mirrored the rise in performance standard. Thus, in both cases many students seemed to be able to fully adapt to the higher stakes or standards. In our view, the size of the *sufficient motivation class* (66% of all students) underscores students' sensitivity to the assessments. Thus, in order to explain students' remarkable sensitivity to assessments, in our final study we developed a theoretical perspective on student motivation in higher education.

Explaining Students' Sensitivity to Assessments

In **chapter 6** we presented a theoretical perspective on motivation in higher education that describes the reasons for students' focus on assessments, as well as the risks of that focus. More specifically, we propose that the degree of alignment between the curricular objectives and assessments plays a pivotal role in student motivation. When there is perfect alignment, all objectives have an equitable probability of being assessed, and thus all learning has an equitable contribution towards assessment performance. As a consequence, it will not matter whether a student is motivated to learn or motivated to perform: learning is a prerequisite for performance, and therefore a student should show the same learning behaviour with both types of motivation.

However, the assessment will likely not be perfectly aligned with the curricular objectives. Cohen-Schotanus (1999) explains that both cognitive (i.e. contents of learning) and operant (i.e. amount of learning) aspects of learning are affected by assessment, and we extended this distinction to identify potential sources of cognitive and operant misalignment. As a result of misalignment, some of the objectives will not need to be achieved in order to pass assessments, resulting in a distinction between assessed objectives and unassessed objectives. As only the assessed objectives benefit students' assessment performance, students are only rewarded for putting effort in assessed objectives. In addition, investing effort in unassessed objectives is implicitly discouraged, as this investment reduces the available time and energy for assessed objectives. Therefore, in terms of probability of good assessment performance, the motivation to perform is more adaptive than the motivation to learn. In other words, students who are motivated to perform have a motivation that fits in a misaligned curriculum.

This curricular fit view on motivation has several implications for educators. The adverse effects of misaligned assessments are not to be underestimated, especially in case of high stakes. Therefore, assessment practices may need to be improved in several ways. Firstly, educators should strive to optimise alignment between assessments and curricular objectives. When assessments are aligned with the objectives, the right learning behaviour is rewarded. Secondly, educators' awareness of the motivational consequences of misaligned assessments needs to improve. Thirdly, assessors need to critically review whether assessments are aimed at weeding out the poor performers, or at stimulating learners to unleash their full potential. And fourthly, educators should cease to consider the assessment as the finish line of the subject, by making exam reviews a more fundamental part of curricula. Finally, when it is impossible to align assessments with curricular objectives, assessments can have detrimental effects on student learning. In that case, educators should consider abandoning assessments, or at least find ways to lower the importance of assessments. In addition to improving assessment practices, educators could make it harder for students to be strategic in allocating their efforts towards the assessed objectives. Therefore, less information about the mode of assessment, as well as about the performance standard, could be given prior to the assessment.

Thus, in chapter 6 we have presented a curricular fit perspective on motivation in higher education, by which we explain students' remarkable sensitivity to the assessment policy observed in the previous chapters: in a misaligned curriculum, the motivation to perform is more adaptive than the motivation to learn, in order to maximise probabilities of graduation. Due to students' sensitivity to assessment, alignment of assessments with curricular objectives (*path a*) becomes a necessary condition for curricula to drive learning in the right direction. An important strength of adopting a curricular fit perspective on motivation, is the resulting focus on aspects of motivation that can be improved by our curricula. We will now discuss the main strengths and limitations of this dissertation, followed by the implications of the findings of this dissertation for educators as well as for future research.

Strengths and limitations

The most important strength of this dissertation is the natural setting in which the investigations were conducted. The assessment policy change at EUR created a unique natural quasi-experiment in which we could compare students under different assessment policies. In our view a natural quasi-experiment is the best option to investigate student motivation, self-regulation, performance and progress under different assessment policies, for two reasons. Firstly, the importance that graduating has for students cannot be induced in an experimental setting. Secondly, it would be unethical to randomly divide a cohort of students into two groups who will have to learn and perform under a different assessment policy.

Additionally, as the policy change was introduced in all course programmes of EUR, and in different ways, we were able to investigate students from various course programmes, and under different assessment policy changes. This for instance created a unique opportunity to investigate student progress differences in three different course programmes in the study in chapter 2. To the best of our knowledge no studies have empirically investigated comparable higher education assessment policy changes, both in terms of the numbers of students involved, and regarding the scope of the changes. Therefore, our investigations added important knowledge to the field of research into student learning under different assessment policies in higher education.

However, the major strength of our studies may also be the major weakness, as the natural setting created three main limitations. Firstly, we cannot draw any causal conclusions based on our investigations, as this research was not experimental. Secondly, this PhD-project started in April 2015, at which time there would be no new first-year cohorts starting under the old assessment policy. This timing posed a challenge regarding the data collection of motivational and self-regulatory constructs: we could not collect data in the old assessment policy anymore. Therefore, for the study reported in chapter 3 we used data that had been collected for a previous study in 2008 and 2009 (Stegers-Jager et al., 2012) as old policy data, and in the investigation described in chapter 4 we used a single course that was a second-year course for psychology and a third-year course for education and child studies (ECS). To remedy threats to the internal validity of the studies, in both studies we performed several checks to ensure the validity of our conclusions.

A third limitation due to the natural setting was that all faculties except business administration made several simultaneous changes to the assessment policy, and

these changes differed per faculty. These simultaneous changes made it impossible to disentangle the separate influences of the elements of the assessment policies under investigation: the stakes, the performance standard, and the resit standard. Additionally, the different changes per faculty made comparisons between faculties less meaningful.

A final limitation concerns the fact that we had to rely on self-report instruments to measure motivation and self-regulation. Self-reports add another layer of uncertainty to the results, as students may have been inclined to give socially desirable answers. However, as explicated in the introduction to this dissertation, we believe self-reports were the preferable way to answer our research questions.

Implications & Directions for future research

This dissertation has several implications for educators and educational policy makers, as well as for future research into assessment policies in higher education. Most importantly, our results imply that the stakes, performance standard and resit standard of assessment policies matter for student learning in several ways. In the following, I will differentiate between what 'we' found (i.e. me, my supervisors and other co-authors), and what 'I' think, as some remarks are my personal reflections based on five years of research into higher education.

Progress and Performance Under Different Assessment Policies

In line with the literature on stakes (Cole & Osterlind, 2008; Sundre & Kitsantas, 2004; Wolf & Smith, 1995) and performance standards (Elikai & Schuhmann, 2010; Johnson & Beck, 1988), we found significant associations of assessment policies with academic progress and academic performance. The significantly faster progress we observed under higher stakes policies saves substantial amounts of energy and time for students and educators. A first consequence is that students complete the first year faster in case of higher stakes. Another consequence of the higher stakes is that the academic dismissal decision is taken after one year instead of two. Thus, as long as the academic dismissal decisions are accurate, both progressing and dismissed students as well as educators benefit in terms of time and energy investment. However, we realise that not all dismissed students may experience their dismissal as a benefit. And of course, based on our findings we cannot draw conclusions on the accuracy of the progress/ academic dismissal decision of the different assessment policies.

Generally, we observed better performance under higher stakes, higher performance standards policies. Thus, in addition to saving time and money, the assessment policy can serve to stimulate student performance. Under the condition that assessments are aligned with the curricular objectives, assessment policies may be an efficient way to improve students' learning.

However, the different conclusions regarding progress and performance in the course programmes of business administration, medicine and psychology underline the importance of a contextualised understanding of the consequences of changes to the assessment policy. For instance, in medicine the stakes in the new assessment policy may have been too high, given the strongly increased early dropout. As a consequence, the medical faculty decided to lower the stakes of the assessment policy three years after the introduction of the changes that were investigated in this dissertation. Thus, raising the stakes can be an efficient way to accelerate academic progress, but the stakes should be achievable. There are several issues to discuss with regards to higher progress and performance under assessment policies with higher stakes and higher performance standards: alignment, underrepresented students, and whether progress is the right goal.

Are the Assessments Sufficiently Aligned?

It is the alignment between assessments and curricular objectives that determines whether high performance is related with good learning. Thus, an important implication of this dissertation is that educators need to optimise alignment, particularly in high stakes assessments. However, a preceding, more fundamental question about alignment of assessments and objectives of education is: is all learning assessable? When considering all the implicit, unobservable aspects of my own learning, I am concerned that the answer may be 'no'. If the answer is indeed 'no', our educational system has a major problem: we are only rewarding the assessable learning, thereby discouraging students from learning what is not assessable (see chapter 6 for an elaborate explanation). This problem is particularly relevant when a sizeable proportion of learning is not assessable. Then, we would either have to stop assessing, or drastically lower the stakes of the assessments. Of course, I realise that not assessing, or even ceasing to attach consequences to students' performance may seem far-fetched. However, in my opinion a necessary requirement for assessment is that it should not obstruct the learning it was intended to monitor in the first place.

Given the stakes attached to assessments in higher education, the implicit answer of most higher education institutions to the question "is all learning assessable?" seems to be 'yes'. Assuming that all learning is assessable, the crucial step is to ensure alignment

of assessments with curricular objectives. Aligned assessments require ample time and energy for educators. A crucial first step would be to improve the amount of time dedicated to training educators' general assessment skills as well as giving educators the time to develop and administer good assessments. For instance, taking the time for in-depth discussions with students requires larger investments than multiple-choice assessments. If the alignment is insufficient, students are implicitly pressured to refrain from learning unassessed objectives, and thus learning suffers. In the long run, society does not benefit so much from good grades; it is good learning that will truly increase the quality of our graduates. Thus, boosting performance on misaligned assessment is a form of short-term thinking that will benefit no one.

What About Underrepresented Students in Higher Education?

In addition to concerns about alignment, underrepresented students deserve particular attention. By underrepresented, I mean students from an ethnic minority background, students from low socioeconomic backgrounds, or first-generation higher education students (i.e. the first from their family to attend higher education). As Berliner (2011) argues from a teacher's perspective, misaligned high-stakes tests may be particularly harmful for these students, as raising the stakes for assessed learning will pressure teachers to refrain from educating the unassessed learning. As we have theoretically substantiated in chapter 6, students experience this same pressure. However, as Berliner (2011) explains, this unassessed learning is potentially crucial to understand a later subject that builds on this learning. And compared with traditional higher education students (i.e. ethnic majority, high socioeconomic background, later generation), the social and cultural capital (Bourdieu, 1986) of underrepresented students shows less resemblance with the social and cultural characteristics of the educational system, throughout students' entire educational career. Therefore, underrepresented students are likely to have relatively less opportunities of learning unassessed skills and knowledge in their private lives. Consequently, underrepresented students will underperform in a later educational stage. Thus, critically paraphrased, the focus on (misaligned) assessments may contribute to the reproduction of societal inequalities.

Apart from misalignment, higher stakes may already be inequitable for underrepresented students. Sometimes circumstances prevent students from showing their potential within the first academic year. Especially underrepresented students may need more time to adapt. A policy report showed that the progress differences under EUR's new versus old assessment policies were comparable in different subgroups of students, based on ethnic background (Baars et al., 2015). However, this report was only descriptive, and did not compare between different levels of socioeconomic

status and first-or-later generation students in higher education. Therefore, future research needs to further clarify which students may be overrepresented among the dropouts under higher-stakes and lower-stakes assessment policies, and why possible overrepresentation exists. In addition, ways to make assessment policies provide more equitable opportunities for all students deserve scrutiny. For instance, allowing some form of compensation between grades is one way to introduce a form of flexibility into the assessment policy, that may help to create equitability.

A continuous difficulty in striving for equitable opportunities for all students is the fact that the difficulty and the attainability of a course programme are hard to separate. This separation can best be understood through an analogy: passing all the assessments in order to graduate from higher education, is like passing all the hurdles of a hurdle race in order to finish. Runners only have a fair chance to show their true potential when there are no potholes and bumps on the hurdle track. In my view, the goal of educators should be to keep the hurdles high (i.e. difficulty), while making sure to remove all the potholes and bumps on the track (i.e. attainability). For instance, assessments should be sufficiently difficult, but only assess appropriate knowledge or skills. Thus, performance on a statistics assessment should depend on students' understanding of statistics, but as little as possible on their level of advanced language comprehension skills. However, when students do not finish (i.e. dropout), in higher education it is often impossible to tell whether this is due to the hurdles being too high, or an unfair consequence of the bumps and potholes. Hopefully, educators will continuously monitor dropouts, in order to make sure students drop out because curricula are too difficult, rather than unattainable.

Is Fast Progress the Right Goal?

As described in chapter 1, the change in assessment policy at EUR was a response to economic and political forces that made satisfactory academic progress a key condition for higher education institutions' healthy financial status. If, for the sake of discussion, we assume that the alignment is perfect, and that performance thus truly reflects achieving the curricular objectives, the question remains whether accelerating academic progress is the right goal for higher education institutions. Learning takes time, and some learners take more time than others. It is up to educators to decide which rate of progress is a suitable goal in their curricula, while keeping in mind that fast learning does not equal good learning.

If academic progress rates are not deemed a good assessment of the true objectives of higher education, the position of higher education institutions in society is analogous to that of a student in a misaligned curriculum: the assessment, i.e. progress rate, is

misaligned with the objectives. Thereby, students as well as educators may experience conflict between their own goals and the goals that are rewarded by their context, i.e. the curriculum or the society, respectively. Consequently, educators may find themselves in a position where letting students pass or progress who do not have the required level of knowledge and/or skills, financially benefits their higher education institution. Therefore, to me it seems inadvisable to uphold a reward structure that financially incentivises higher education institutions to boost academic progress rates.

In conclusion, the significant differences we observed in terms of academic progress and performance under different assessment policies can be interpreted both positively and negatively. Crucial questions to answer are whether the assessments are sufficiently aligned with the curricular objectives, what happens to underrepresented students, and whether fast progress is the right goal for higher education institutions.

Motivation and Self-Regulation Under Different Assessment Policies

As the literature on motivation and self-regulation under different characteristics of assessment policies is scarce (Sundre & Kitsantas, 2004; Sungur, 2007), this dissertation fills an important gap in the literature. Overall, in the studies described in chapters 3 and 4, we observed that both medical students and social science students had higher competence beliefs (i.e. self-efficacy), valued their learning more (i.e. task value), and managed their time better under assessment policies with higher stakes and higher performance standards. Thus, in general it seems that raising the stakes and performance standards is an efficient way to improve motivation and self-regulation. However, our results also raise questions concerning possible negative effects for motivation, the best measurement of motivation and alignment, and what level of motivation is 'sufficient'.

What are Possible Negative Consequences of More Difficult Assessment Policies?

Despite our generally positive findings regarding motivation and self-regulation, there are causes for concern as well. Most importantly, the negative consequences of misalignment for motivation and learning may increase when the stakes are raised. Additionally, results of chapter 4 indicate that students under a more difficult assessment policy experience more test anxiety. Furthermore, a recent study on medical students indicated that particularly the female students experience more stress in a higher-stakes assessment policy (Stegers-Jager et al., 2020). Besides these findings on test anxiety and stress, we have no data to compare between the different assessment policies concerning possible negative effects of more difficult assessment policies in terms of students' well-being.

Furthermore, long-term negative consequences that more difficult assessment policies could have for motivation should be investigated in future studies. More specifically, we have observed that setting a difficult, specific goal for students (i.e. obtain all first-year credits within one year) seems to motivate them to progress faster. However, what if setting personal goals is an important skill that students should learn during their education? In that case, setting the goal for students deprives them of the opportunity to develop their own goal-setting skills.

Another cause for concern, are several motivational researchers' worries about the negative effects that extrinsic motivators, such as assessments, may have on students' intrinsic motivation (Deci et al., 1999; Harlen & Crick, 2003). One of our motivational constructs was task value, i.e. the extent to which a student finds the material interesting and worth learning (Credé & Phillips, 2011). Task value is considered a representative measure of academic intrinsic motivation (Richardson, Abraham, & Bond 2012). Therefore, from our findings on motivation and self-regulation, especially the higher task value under higher stakes, higher performance standard assessment policies was remarkable.

We can think of three, largely compatible, explanations for this higher task value. Firstly, while recognizing the importance of intrinsic motivation, Hidi and Harackiewicz (2000) argue that external motivators can have powerful additive value in motivating students. For instance, sometimes certain crucial topics (e.g. methods and statistics) are not the most interesting, nor intrinsically motivating topics for students. In those instances, assessments can help to motivate students to still sufficiently master those topics. Thus, instead of viewing intrinsic and extrinsic motivators as antagonistic, it may be the case that both serve different roles in performance (Cerasoli et al., 2014). Secondly, in chapter 6 we have argued that only assessments that are misaligned with curricular objectives will have a negative effect on student motivation. Then perhaps, setting higher grade goals and/or progress goals would decrease task value on misaligned assessments, yet increase task value on aligned assessments. Thus, the higher task value under the higher stakes, higher performance standard assessment policies may reflect alignment of assessments with the objectives. Thirdly, only the stakes and standards changed, not the number of assessments. Perhaps these difficult and specific new stakes and standards were more motivating, at least for students who considered these stakes and standards attainable (Locke & Latham, 2002).

This third explanation for the higher task value relates to another possible negative consequence of more difficult policies: the one-year timeframe of the new assessment policy was not feasible for a portion of students. As we have noticed in chapter 2, not all

medical and business administration students seemed to be able to complete the first academic year in one year instead of two. Thus, although students generally reported higher motivation under the new policy, particularly (early) dropouts may have been demotivated by the stakes and performance standard of the new policy. Especially the motivational processes of these dropouts require further scrutiny: why were these students demotivated, and were their expectations of success accurate?

How to Measure Motivation?

In chapter 6, we theoretically substantiated the motivational importance of grades to students, which is supported by the importance of grade goals and performance self-efficacy as predictors of academic performance (Richardson et al., 2012; Schneider & Preckel, 2017). In addition, aimed and minimum grade goals, as well as performance self-efficacy, are all one-item measures. Therefore, grade goals and performance self-efficacy have the potential to be highly efficient and relevant motivational constructs for both educators and motivational researchers.

In fact, in case of perfect alignment, students' grade goals indicate how much students want to learn of the subject's objectives. As an educator, it is a very quick way to learn more about the goals of your students. Knowing students' goals is important in order to support students in attaining those goals, as well as in helping students to possibly adjust their goals to the right level of challenge. For researchers, grade goals can be used to better understand the causes of performance, as well as the consequences that curriculum changes have for motivation.

In case of misalignment, in addition to grade goals, another relevant question to ask students is: 'To what extent do you want to do the unassessed learning?'. It would be interesting to investigate how the answer to this question is associated with performance, as the unassessed learning lowers the available time and energy for performing the assessed learning. However, in addition to this question, the accuracy of students' expectations of misalignment would need to be measured as well, as accurate expectations are a necessary condition in order to be able to predict which learning will or will not be assessed. A necessary condition in order to be able to determine the accuracy of these expectations, is that the alignment is known as well.

How to Measure Alignment?

Based on the findings of this dissertation, I have concluded that assessment policies can be an effective source of improved performance. The degree of alignment between assessments and curricular objectives determines the extent to which this

improved performance reflects improved learning. Thus, determining the degree of alignment between assessments and objectives in contemporary higher education is an important direction for future research as well. However, measuring alignment is easier said than done. Only someone with a full overview of an entire curriculum and all the curricular subjects can potentially assess curricular alignment. Not many teachers in higher education will have a full overview of the curriculum in which they teach. Therefore, perhaps the best option is to ask recent graduates to assess the curricular alignment, as graduates have just completed the entire curriculum. Of course, graduates' evaluations of alignment would need to be supplemented with educators' evaluations of alignment. Educators would include teachers, course programme directors, and assessment experts. A comparison of students' and educators' evaluations would be valuable, as this comparison could shed light on differences in perceptions of the curriculum. Concretely, I expect many teachers to overestimate the alignment between the assessment and objectives of their subject, and many course programme directors to overestimate the alignment of assessments and curricular objectives.

Evaluating alignment most likely needs to be performed per curricular objective, and will thus be highly context-specific. Thus, assessing alignment will require substantial investments of time and effort, yet only directly benefit one particular curriculum. However, given the detrimental consequences of misalignment (e.g. discouraging unassessed learning), in my view the benefits outweigh the investments due to improvements in student learning. Thus, more concretely, future research should address the question: what curricular objectives are underrepresented in our assessments?

In addition to these direct benefits for specific curricula, assessing alignment has several indirect benefits as well. Firstly, investigating the alignment in individual curricula will elucidate the current state of affairs: how well are the assessments and curricular objectives aligned in contemporary higher education? Secondly, these investigations may clarify best-practices regarding alignment, and inspire others to improve alignment as well. And thirdly, these investigations may clarify under which conditions alignment is unfeasible, for instance because some learning is not assessable. Or in other words, these investigations will help to answer: which learning is harder, or impossible to assess?

What Level of Motivation is Sufficient?

In chapter 5, we observed that about two thirds (66%) of students, i.e. the students in the *sufficient motivation class,* seems particularly sensitive to the performance

standard in terms of their grade goals and performance self-efficacy. It is important to note that this was the case in a particular sample of social science students in a specific curriculum. More research is needed in order to elucidate (the limits to) the generalisability of this percentage of students with a motivational focus on the minimum performance standard.

I expect that a portion of these students could be aiming for better performance than the minimum level needed to pass. Therefore, my personal inclination is to find out how to make these students aim for higher grades. However, perhaps my inclination is misplaced, as aiming for the minimum grade needed to pass is not necessarily troublesome. Saving energy on assessment preparation can leave room for students to pursue their own interests, first and foremost within the curriculum. And in case of misalignment, the learning process of a student may lead to better long-term outcomes in case of lower grade goals. Therefore, instead of raising students' goals, future research will first need to show what different reasons exist for students to have a strong motivational focus on the minimum performance standard.

Importantly, educators are responsible for determining a suitable minimum performance standard. Thus, instead of motivating students to set higher goals, educators should make sure the performance standard is accurate in discerning between students who should pass, and students who should fail. In chapter 4 we observed that none of the students had minimum grade goals below the performance standard, which denotes that all students at least intended to pass. This observation further underlines the importance of choosing an optimal performance standard, at which educators are confident that decisions about students' level of learning are accurate. In addition to ensuring decision accuracy, educators have the responsibility to determine performance standards with beneficial motivational consequences.

In sum, in addition to progress and performance, motivation and self-regulation are associated with assessment policies. Therefore, assessments do not merely establish students' level of performance: assessments are a fundamental curricular element that serve a crucial function in learning. In this dissertation, we have attempted to demonstrate this pivotal role of assessment in student learning, both empirically and theoretically.

Stakes, Performance Standards, and Resit Standards of the Assessment Policy

The stakes, performance standards and resit standards of assessment policies jointly affect students' learning. However, there is still a lot to be learned by investigating

what happens to student learning in case of isolated changes to either the stakes, performance standards or resit standards. In the end, educators will want to know: what are 'the right' stakes, performance standard, and resit standard? Although this question needs to be answered context-specifically, our results give some food for thought.

What are the 'Right' Stakes, Performance Standards, and Resit Standards?
We have already discussed the possible negative consequences of high stakes. Putting these negative consequences aside, in general, based on our findings we conclude that raising the stakes seems to be an efficient way to accelerate academic progress. As, none of EUR's course programmes only changed the performance standard or resit standard, based on our results we can only speculate on what performance standards and resit standards are most beneficial for student learning. Regarding the resit standard, our findings are in line with previous findings: Offering students opportunities for resits may induce a certain reliance on second chances (Scott, 2012), and result in lower effort (Nijenkamp et al., 2016) as well as lower performance (Grabe, 1994) on the first attempt. As resits also offer an opportunity to unjustly pass an assessment by chance (Yocarini et al., 2018), traditional resit policies require scrutiny in future research. Context-specific answers need to be sought to questions about the best number of resits, both in terms of decision accuracy and student effort and/or performance. Overall, I would advise educators to be reluctant in giving many resit opportunities to students without personal circumstances.

Regarding the performance standards, our findings are also consistent with the available literature: raising the performance standard can be a powerful source of improved performance (Elikai & Schuhmann, 2010; Johnson & Beck, 1988), and allowing compensation between grades will result in different selections for progress (Yocarini et al., 2018). Due to these consequences for performance and selection, performance standards should be critically examined. For instance, the traditional Dutch performance standard in which each individual assessment needs to be passed with at least a 5.5 on a 10-point scale, will likely not be the best performance standard in every curriculum. For instance, compensatory standards should result in better progress decisions in case of highly correlated subjects (Yocarini et al., 2018).

Additionally, a 5.5 performance standard corresponds to 50-60% mastery. The question is whether this percentage of mastery should be the same in different school levels, between different curricula, or across courses within the same curriculum. I suspect that the 5.5 standard is often chosen due to traditions instead of careful considerations of the consequences for performance and selection. The question is whether there is confidence that students with the lowest passing grade indeed have sufficiently

mastered the objectives. This dissertation will hopefully raise educators' awareness of the motivational consequences of their performance standard.

The question remains what will happen to academic performance when the performance standards are raised even further than was currently the case. Based on our and previous findings (Lindo et al., 2010), in case of still higher standards we would generally expect higher drop-out yet improved performance of the remaining students. However, rather than raising stakes, performance standards and/or resit standards even further, a more provocative option is to let the stakes and standards be unknown a priori, to students, or even educators.

Should the Standards be Known A Priori?

Given students' sensitivity to the stakes and standards we observed in the studies in chapters 2-5, an interesting option is not to communicate the stakes and standards to the students prior to the assessment. Concretely, that would mean that students do not know how many credits to obtain in order to avoid dismissal, what grades are sufficient to obtain credits, or how many resit opportunities they will get. The question is what students would aim for when there is no minimum to aim for. Given our observation that about two thirds of students set grade goals near the performance standard, I would expect many of these students to aim higher when the minimum is unknown. Consequently, although we suspect this non-communication will not be appreciated by many students in the short term, and may cause stress, it could be a powerful push for improved performance. Future research should examine whether students are inclined to perform better when the stakes and standards are not provided before the assessments.

Additionally, even a situation in which the educators do not know the stakes and standards is possible; that is, when the stakes, performance standards, and/or resit standards are not determined a priori. Instead, the stakes and standards could be determined a posteriori, and possibly per individual student, by an expert panel (Van der Vleuten et al., 2012). Letting an expert panel decide on individual students' progress, sufficient performance, or number of resits may appear more subjective than setting a priori standards that are the same for each student, but subjectivity does not equal arbitrariness.

In fact, a posteriori stakes and standards can take at least three sources of information into account, and thereby result in more accurate decisions about students' performance and progress, and/or lead to increased motivation. Firstly, as the difficulty of assessments fluctuates, answering 50% correctly on two different assessments of

the same subject does not necessarily imply the same level of mastery. Therefore, taking information on the difficulty of an assessment into account when determining the performance standard may improve the decision accuracy. For instance, when even the highest-scoring students have low scores, this may indicate that an assessment was very difficult. Consequently, using the highest-scoring students as a reference point in determining the performance standard on individual assessments (see Cohen-Schotanus & Vleuten, 2010), seems to improve the estimation of students' ability (Yocarini, 2019). Thus, a posteriori performance standards improve the decision accuracy for individual assessments. Extrapolating these conclusions with regard to decisions about student progress implies that determining the standard for a combination of multiple assessments should also be more accurate when taking assessment difficulty into account.

Secondly, the quality of assessments can be taken into account in case of a posteriori standards. For instance, suppose that out of all first-year assessments, one assessment turns out to be highly unreliable. Basing students' progress decision on this assessment may be problematic, hence the weighting of this assessment in the progress decision could be adapted. Thus, the accuracy of decisions about progress would increase by considering the quality of assessments.

Thirdly, a posteriori standards could make it possible to consider information about individual students. Students may have had unforeseen circumstances, or a difficult adaptation. Therefore, instead of telling students whether they have passed the bar, students could be involved in the discussion of where the bar should be placed. To promote ownership of students' goals, students could even set their own grade goals and progress goals at the start of the year, reflect on these goals during the year, and discuss why the goals were or were not attained. Of course, students would have to be guided in setting appropriate goals, as well as in reflecting on their process towards attaining those goals. For instance, educators would need to monitor that the goals are not too low for students in the *sufficient motivation class*, and perhaps also not too high for students in the *high motivation class.* Future research would need to show what goals students would set then, and how goal ownership affects motivation and performance.

Due to the possibility of taking difficulty, quality, and individual students into account, a posteriori stakes and standards may well be less arbitrary than a priori stakes and standards in terms of decision accuracy. Additionally, a posteriori stakes and standards could also serve an important motivational function, by making it more difficult for students to only learn what is needed to minimally pass. If the stakes and standards

do not exist yet, there is also no minimum to aim for. Future research is necessary to find out whether decision accuracy and learning indeed improve when the stakes and standards are unknown to students, or set a posteriori.

A major difficulty of a posteriori stakes and standards is the requirement of expert panels to determine those stakes and standards. As determining the stakes and standards requires knowledge about the content of the curriculum as well as psychometrics, these panels would need to consist of both content experts and psychometricians. Experts will cost money, and good assessment takes time and thus money as well. However, I hope to have substantiated that investment of time and money is a bare necessity given the current importance of assessments in our educational system. I expect that increased assessment-expertise will have another important advantage: assessment experts are less likely to overestimate the accuracy of grades. Even for well-aligned, reliable and valid assessments, the grade attached to a student's performance will be a rough estimation of a student's knowledge and skills, rather than a perfectly accurate representation.

Conclusion

The most important conclusion of this dissertation is that higher education students' academic progress, academic performance, motivation and self-regulation seem sensitive to the assessment policy. Although the extent of this sensitivity was remarkable, the sensitivity makes sense: students aim to graduate, and the assessment policy determines what is required to graduate. Thus, it is adaptive for students to be sensitive to the assessment policy, in order to raise the probability of graduating. Therefore, assessments are an important motivational tool for those who devise the curriculum and adapting the assessment policy can be a highly effective source of improved performance. The degree of alignment between the curricular objectives and assessments determines the degree to which improved performance reflects improved learning as intended by the curricular objectives. Thus, under the condition that the importance of assessments in our curricula is reflected in the amount of time and energy devoted to ensuring aligned assessments, grades reward the right learning behaviour. Then, students' sensitivity to the assessment policy can be harnessed to improve learning. Learning more will require more effort, yet will benefit students in the long run. Thus, challenging assessments can be a didactical act of love.

REFERENCES

Akgun, S., & Ciarrochi, J. (2003). Learned resourcefulness moderates the relationship between academic stress and academic performance. *Educational Psychology*, *23*(3), 287–294. https://doi.org/10.1080/0144341032000060129

Al-Kadri, H. M., Al-Moamary, M. S., Roberts, C., & Van der Vleuten, C. P. M. (2012). Exploring assessment factors contributing to students' study strategies: Literature review. *Medical Teacher*, *34 Suppl 1*, S42-50. https://doi.org/10.3109/0142159X.2012.656756

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct*. https://www.apa.org/ethics/code/

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, *84*(3), 261–271.

Anderson, L. W. (2002). Curricular alignment: A re-examination. *Theory Into Practice*, *41*(4), 255–260. https://doi.org/10.1207/s15430421tip4104_9

Arbuckle, J. L. (2013). *AMOS* (Version 22.0) [Computer software]. IBM SPSS.

Arnold, I. J. M. (2012). De BSA-norm: Minimumeis of streef- waarde? *Tijdschrift voor Hoger Onderwijs & Management*, *3*, 4–8.

Arnold, I. J. M. (2015). The effectiveness of academic dismissal policies in Dutch university education: An empirical investigation. *Studies in Higher Education*, *40*(6), 1068–1084. https://doi.org/10.1080/03075079.2013.858684

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, *64*(6, Pt.1), 359–372. https://doi.org/10.1037/h0043445

Attewell, P., Heil, S., & Reisel, L. (2011). Competing explanations of undergraduate noncompletion. *American Educational Research Journal*, *48*(3), 536–559. https://doi.org/10.3102/0002831210392018

Baars, G., van Wensveen, P., & Hermus, P. (2015). *Doorwerking regeling "Nominaal = Normaal" aan de Erasmus Universiteit Rotterdam*. Erasmus Universiteit Rotterdam/Risbo. https://www.risbo.org/rapporten/Rapportage%20Onderzoek%20Doorwerking%20Nominaal%20is%20Normaal%20aan%20de%20EUR.pdf

Baars, G., Wolff, R., Godor, B., & Hermus, P. (2013). *Uitgebreide pilot "Nominaal = Normaal" aan de Erasmus Universiteit Rotterdam*. Erasmus Universiteit Rotterdam/Risbo. https://www.risbo.org/rapporten/Onderzoeksrapport%20Risbo%20Uitgebreide%20Pilot%20Nominaal%20is%20Normaal.pdf

Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & Van der Vleuten, C. P. M. (2007). Evaluating assessment quality in competence-based education: A qualitative comparison of two frameworks. *Educational Research Review*, *2*(2), 114–129. https://doi.org/10.1016/j.edurev.2007.06.001

Backer, D. I., & Lewis, T. E. (2015). Retaking the test. *Educational Studies*, *51*(3), 193–208. https://doi.org/10.1080/00131946.2015.1033524

Baeten, M., Kyndt, E., Struyven, K., & Dochy, F. (2010). Using student-centred learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness. *Educational Research Review*, *5*(3), 243–260. https://doi.org/10.1016/j.edurev.2010.06.001

Bandaranayake, R. C. (2008). Setting and maintaining standards in multiple choice examinations: AMEE Guide No. 37. *Medical Teacher*, *30*(9–10), 836–845. https://doi.org/10.1080/01421590802402247

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*(2), 122–147. https://doi.org/10.1037/0003-066X.37.2.122

Becker, H. S., Geer, B., & Hughes, E. C. (1968). *Making the grade: The academic side of college life*. Wiley & Sons.

Berliner, D. (2011). Rational responses to high stakes testing: The case of curriculum narrowing and the harm that follows. *Cambridge Journal of Education*, *41*(3), 287–302. https://doi.org/10.1080/0305764X.2011.607151

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, *32*(3), 347–364. https://doi.org/10.1007/BF00138871

Blom, S., & Severiens, S. (2008). Engagement in self-regulated deep learning of successful immigrant and non-immigrant students in inner city schools. *European Journal of Psychology of Education*, *23*(1), 41–58. https://doi.org/10.1007/BF03173139

Boatright-Horowitz, S. L., & Arruda, C. (2013). College students' categorical perceptions of grades: It's simply 'good' vs. 'bad'. *Assessment & Evaluation in Higher Education*, *38*(3), 253–259. https://doi.org/10.1080/02602938.2011.618877

Boer, H. F. de, Jongbloed, B. W. A., Benneworth, P. S., Cremonini, L., Kolster, R., Kottmann, A., Lemmens-Krug, K., & Vossensteyn, J. J. (2015). *Performance-based funding and performance agreements in fourteen higher education systems*. https://research.utwente.nl/en/publications/performance-based-funding-and-performance-agreements-in-fourteen-

Boud, D. (1995). Assessment and learning: Contradictory or complementary? In P. Knight (Ed.), *Assessment for Learning in Higher Education*. (pp. 35–48). Kogan Page.

Boud, D., Dawson, P., Bearman, M., Bennett, S., Joughin, G., & Molloy, E. (2018). Reframing assessment research: Through a practice perspective. *Studies in Higher Education*, *43*(7), 1107–1118. https://doi.org/10.1080/03075079.2016.1202913

Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of Theory and Research for the Sociology of Education* (pp. 241–258). Greenwood. https://doi.org/10.1002/9780470755679.ch15

Britton, B. K., & Tesser, A. (1991). Effects of time-management practices on college grades. *Journal of Educational Psychology*, *83*(3), 405–410. https://doi.org/10.1037/0022-0663.83.3.405

Broekkamp, H., & Van Hout-Wolters, B. H. A. M. (2006). Students' adaptation of study strategies when preparing for classroom tests. *Educational Psychology Review*, *19*(4), 401. https://doi.org/10.1007/s10648-006-9025-0

Byrne, B. M. (2013). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Routledge.

Campbell, D. T. (1976). *Assessing the impact of planned social change. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=8083EC66C38CAA01FC82574D7D06C37C?doi=10.1.1.170.6988&rep=rep1&type=pdf*.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, *13*(2), 195–212. https://doi.org/10.1007/BF01246098

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin*, *140*(4), 980–1008. https://doi.org/10.1037/a0035661

Chester, M. D. (2003). Multiple measures and high-stakes decisions: A framework for combining measures. *Educational Measurement: Issues and Practice*, *22*(2), 32–41. https://doi.org/10.1111/j.1745-3992.2003.tb00126.x

Cilliers, F. J., Schuwirth, L. W., Adendorff, H. J., Herman, N., & van der Vleuten, C. P. M. (2010). The mechanism of impact of summative assessment on medical students' learning. *Advances in Health Sciences Education*, *15*(5), 695–715. https://doi.org/10.1007/s10459-010-9232-9

Cilliers, F. J., Schuwirth, L. W. T., Herman, N., Adendorff, H. J., & Van der Vleuten, C. P. M. (2011). A model of the pre-assessment learning effects of summative assessment in medical education. *Advances in Health Sciences Education*, *17*(1), 39–53. https://doi.org/10.1007/s10459-011-9292-5

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cohen, S. A. (1987). Instructional alignment: Searching for a magic bullet. *Educational Researcher*, *16*(8), 16–20. https://doi.org/10.3102/0013189X016008016

Cohen-Schotanus, J. (1999). Student assessment and examination rules. *Medical Teacher*, *21*(3), 318–321. https://doi.org/10.1080/01421599979626

Cohen-Schotanus, J., & Van der Vleuten, C. P. M. (2010). A standard setting method with the best performing students as point of reference: Practical and affordable. *Medical Teacher*, *32*(2), 154–160. https://doi.org/10.3109/01421590903196979

Cole, J. S., & Osterlind, S. J. (2008). Investigating differences between low-and high-stakes test performance on a general education exam. *The Journal of General Education*, *57*(2), 119–130.

Conley, A. M. (2012). Patterns of motivation beliefs: Combining achievement goal and expectancy-value perspectives. *Journal of Educational Psychology*, *104*(1), 32–47. https://doi.org/10.1037/a0026042

Cook, D. A., & Artino, A. R. (2016). Motivation to learn: An overview of contemporary theories. *Medical Education*, *50*(10), 997–1014. https://doi.org/10.1111/medu.13074

Cook, D. A., Thompson, W. G., & Thomas, K. G. (2011). The Motivated Strategies for Learning Questionnaire: Score validity among medicine residents. *Medical Education*, *45*(12), 1230–1240. https://doi.org/10.1111/j.1365-2923.2011.04077.x

Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology*, *51*(1), 171–200. https://doi.org/10.1146/annurev.psych.51.1.171

Credé, M., & Phillips, L. A. (2011). A meta-analytic review of the Motivated Strategies for Learning Questionnaire. *Learning and Individual Differences*, *21*(4), 337–346. https://doi.org/10.1016/j.lindif.2011.03.002

Credé, M., Roch, S. G., & Kieszczynka, U. M. (2010). Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research*, *80*(2), 272–295. https://doi.org/10.3102/0034654310362998

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302.

De Koning, B. B., Loyens, S. M. M., Rikers, R. M. J. P., Smeets, G., & Van der Molen, H. T. (2014). Impact of binding study advice on study behavior and pre-university education qualification factors in a problem-based psychology bachelor program. *Studies in Higher Education*, *39*(5), 835–847. https://doi.org/10.1080/03075079.2012.754857

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*(6), 627–668. https://doi.org/10.1037/0033-2909.125.6.627

Douglas, K. M., & Mislevy, R. J. (2010). Estimating classification accuracy for complex decision rules based on multiple scores. *Journal of Educational and Behavioral Statistics*, *35*(3), 280–306. https://doi.org/10.3102/1076998609346969

Driessen, E., & Van Der Vleuten, C. P. M. (2000). Matching student assessment to problem-based learning: Lessons from experience in a law faculty. *Studies in Continuing Education*, *22*(2). https://doi.org/10.1080/713695731

Duncan, T. G., & McKeachie, W. J. (2005). The making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist*, *40*(2), 117–128. https://doi.org/10.1207/s15326985ep4002_6

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*(10), 1040.

Eijsvogels, T. M. H., Goorden, R., Van den Bosch, W., & Hopman, M. T. E. (2015). The binding study advice in medical education: A 2-year experience. *Perspectives on Medical Education*, *4*(1), 39–42. https://doi.org/10.1007/s40037-015-0164-1

Elikai, F., & Schuhmann, P. W. (2010). An examination of the impact of grading policies on students' achievement. *Issues in Accounting Education*, *25*(4), 677–693. https://doi.org/10.2308/iace.2010.25.4.677

Entwistle, N. J., & Entwistle, A. (1991). Contrasting forms of understanding for degree examinations: The student experience and its implications. *Higher Education*, *22*(3), 205–227. https://doi.org/10.1007/BF00132288

Erasmus MC. (2014). *Education- and examination regulations medical school N=N bachelor's degree 2014-2015*. https://www.erasmusmc.nl/geneeskunde-cs/docume nten/20142015/OER_totaal_web_2014.pdf

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. SAGE.

Finley, J. R., & Benjamin, A. S. (2012). Adaptive and qualitative changes in encoding strategy with experience: Evidence from the test-expectancy paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 632–652.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist*, *39*(3), 193–202. https://doi.org/10.1037/0003-066X.39.3.193

Gaultney, J. F., & Cann, A. (2001). Grade expectations. *Teaching of Psychology*, *28*(2), 84–87. https://doi.org/10.1207/S15328023TOP2802_01

Gibbs, G., & Simpson, C. (2005). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, *1*, 3–31.

Glaser, R., & Kiecolt-Glaser, J. K. (2005). Stress-induced immune dysfunction: Implications for health. *Nature Reviews Immunology*, *5*(3), 243–251. https://doi.org/10.1038/nri1571

Govaerts, M., & Van der Vleuten, C. P. M. (2013). Validity in work-based assessment: Expanding our horizons. *Medical Education*, *47*(12), 1164–1174. https://doi.org/10.1111/medu.12289

Grabe, M. (1994). Motivational deficiencies when multiple examinations are allowed. *Contemporary Educational Psychology*, *19*(1), 45–52. https://doi.org/10.1006/ceps.1994.1005

Guillaume, D. W., & Khachikian, C. S. (2011). The effect of time-on-task on student grades and grade expectations. *Assessment & Evaluation in Higher Education*, *36*(3), 251–261. https://doi.org/10.1080/02602930903311708

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, *92*(1), 160–170. https://doi.org/10.1037/0022-0663.92.1.160

Harlen, W., & Crick, R. D. (2003). Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice*, *10*(2), 169–207. https://doi.org/10.1080/0969594032000121270

Heeneman, S., Oudkerk Pool, A., Schuwirth, L. W. T., van der Vleuten, C. P. M., & Driessen, E. W. (2015). The impact of programmatic assessment on student learning: Theory versus practice. *Medical Education*, *49*(5), 487–498. https://doi.org/10.1111/medu.12645

Heijne-Penninga, M., Kuks, J. B. M., Hofman, W. H. A., & Cohen-Schotanus, J. (2008). Influence of open- and closed-book tests on medical students' learning approaches. *Medical Education*, *42*(10), 967–974. https://doi.org/10.1111/j.1365-2923.2008.03125.x

Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research*, *70*(2), 151–179. https://doi.org/10.3102/00346543070002151

Hossain, B., & Tsigaris, P. (2015). Are grade expectations rational? A classroom experiment. *Education Economics*, *23*(2), 199–212. https://doi.org/10.1080/09645292.2012.735073

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

IBM SPSS Statistics for Windows (Version 23.0). (2014). [Computer software]. IBM Corp.

IBM SPSS Statistics for Windows (Version 24.0). (2016). [Computer software]. IBM Corp.

Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test…or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review*, *26*(2), 307–329. https://doi.org/10.1007/s10648-013-9248-9

Johnson, B. G., & Beck, H. P. (1988). Strict and lenient grading scales: How do they affect the performance of college students with high and low SAT scores? *Teaching of Psychology*, *15*(3), 127–131. https://doi.org/10.1207/s15328023top1503_4

Jung, T., & Wickrama, K. a. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, *2*(1), 302–317. https://doi.org/10.1111/j.1751-9004.2007.00054.x

Khachikian, C. S., Guillaume, D. W., & Pham, T. K. (2011). Changes in student effort and grade expectation in the course of a term. *European Journal of Engineering Education*, *36*(6), 595–605. https://doi.org/10.1080/03043797.2011.637189

Kickert, R., Meeuwisse, M., Jongerling, J., Stegers-Jager, K. M., Arends, L. R., & Prinzie, P. (2020). Grade goals and performance self-efficacy throughout the first academic year: A latent class approach. *Manuscript Submitted for Publication*.

Kickert, R., Meeuwisse, M., Stegers-Jager, K. M., Koppenol-Gonzalez, G. V., Arends, L. R., & Prinzie, P. (2019). Assessment policies and academic performance within a single course: The role of motivation and self-regulation. *Assessment & Evaluation in Higher Education*, *44*(8), 1177–1190. https://doi.org/10.1080/02602938.2019.1580674

Kickert, R., Stegers-Jager, K. M., Meeuwisse, M., Prinzie, P., & Arends, L. R. (2018). The role of the assessment policy in the relation between learning and performance. Medical Education, 52(3), 324–335. https://doi.org/10.1111/medu.13487

Kitsantas, A. (2002). Test preparation and performance: A self-regulatory analysis. *The Journal of Experimental Education*, *70*(2), 101–113. https://doi.org/10.1080/00220970209599501

Knekta, E. (2017). Are all pupils equally motivated to do their best on all tests? Differences in reported test-taking motivation within and between tests with different stakes. *Scandinavian Journal of Educational Research*, *61*(1), 95–111. https://doi.org/10.1080/00313831.2015.1119723

Komarraju, M., & Nadler, D. (2013). Self-efficacy and academic achievement: Why do implicit beliefs, goals, and effort regulation matter? *Learning and Individual Differences*, *25*, 67–72. https://doi.org/10.1016/j.lindif.2013.01.005

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice*, *41*(4), 212–218.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134.

Lambe, P., & Bristow, D. (2011). Predicting medical student performance from attributes at entry: A latent class analysis. *Medical Education*, *45*(3), 308–316. https://doi.org/10.1111/j.1365-2923.2010.03897.x

Laursen, B., & Hoff, E. (2006). Person-centered and variable-centered approaches to longitudinal data. *Merrill-Palmer Quarterly*, *52*(3), 377–389.

Lindo, J. M., Sanders, N. J., & Oreopoulos, P. (2010). Ability, gender, and performance standards: Evidence from academic probation. *American Economic Journal: Applied Economics*, *2*(2), 95–117. https://doi.org/10.1257/app.2.2.95

Lindquist, E. F. (1951). *Preliminary considerations in objective test construction.pdf: EDU S061A1: Statistical and Psychometric Methods for Educational Measurement (Part I)*. https://canvas.harvard.edu/courses/33644/files/5027562

Linnenbrink-Garcia, L., Wormington, S. V., Snyder, K. E., Riggsbee, J., Perez, T., Ben-Eliyahu, A., & Hill, N. E. (2018). Multiple pathways to success: An examination of integrative motivational profiles among upper elementary and college students. *Journal of Educational Psychology*, *110*(7), 1026–1048. https://doi.org/10.1037/edu0000245

Locke, E. A., & Bryan, J. F. (1968). Grade goals as determinants of academic achievement. *The Journal of General Psychology*, *79*(2), 217–228. https://doi.org/10.1080/00221309.1968.9710469

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*. Prentice-Hall, Inc.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, *57*(9), 705–717. https://doi.org/10.1037/0003-066X.57.9.705

Lundeberg, M. A., & Fox, P. W. (1991). Do laboratory findings on test expectancy generalize to classroom outcomes? *Review of Educational Research*, *61*(1), 94–106. https://doi.org/10.3102/00346543061001094

Maskey, V. (2012). Grade expectation and achievement: Determinants and influential relationships in business courses. *American Journal of Educational Studies*, *5*(1), 71–88.

McDaniel, M. A., Blischak, D. M., & Challis, B. (1994). The effects of test expectancy on processing and memory of prose. *Contemporary Educational Psychology*, *19*(2), 230–248. https://doi.org/10.1006/ceps.1994.1019

McManus, I. C. (1992). Does performance improve when candidates resit a postgraduate examination? *Medical Education*, *26*(2), 157–162. https://doi.org/10.1111/j.1365-2923.1992.tb00142.x

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care*, *44*(11), S69–S77.

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., & Urdan, T. (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS)*. University of Michigan.

Miller, C., & Parlett, M. (1974). *Up to the mark: A study of the examination game*. Society for Research into Higher Education.

Muthén, L. K., & Muthén, B. O. (1998). *Mplus User's Guide* (Seventh edition). Muthén & Muthén.

Newble, D. I., & Jaeger, K. (1983). The effect of assessments and examinations on the learning of medical students. *Medical Education*, *17*(3), 165–171. https://doi.org/10.1111/j.1365-2923.1983.tb00657.x

Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, *91*(3), 328–346. https://doi.org/10.1037/0033-295X.91.3.328

Nijenkamp, R., Nieuwenstein, M. R., Jong, R. de, & Lorist, M. M. (2016). Do resit exams promote lower investments of study time? Theory and data from a laboratory study. *PLOS ONE*, *11*(10), e0161708. https://doi.org/10.1371/journal.pone.0161708

Öhrstedt, M., & Scheja, M. (2018). Targeting efficient studying – first-semester psychology students' experiences. *Educational Research*, *60*(1), 80–96. https://doi.org/10.1080/00131881.2017.1406314

Olivier, J., & Bell, M. L. (2013). Effect sizes for 2×2 contingency tables. *PLoS ONE*, *8*(3), e58777. https://doi.org/10.1371/journal.pone.0058777

Onwuegbuzie, A. J., & Wilson, V. A. (2003). Statistics anxiety: Nature, etiology, antecedents, effects, and treatments–a comprehensive review of the literature. *Teaching in Higher Education*, *8*(2), 195–209. https://doi.org/10.1080/1356251032000052447

Pastor, D. A., Barron, K. E., Miller, B. J., & Davis, S. L. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemporary Educational Psychology*, *32*(1), 8–47. https://doi.org/10.1016/j.cedpsych.2006.10.003

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, *36*(1), 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002

Pell, G., Boursicot, K., & Roberts, T. (2009). The trouble with resits … *Assessment & Evaluation in Higher Education*, *34*(2), 243–251. https://doi.org/10.1080/02602930801955994

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, *95*(4), 667–686. https://doi.org/10.1037/0022-0663.95.4.667

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, *16*(4), 385–407. https://doi.org/10.1007/s10648-004-0006-x

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40. https://doi.org/10.1037/0022-0663.82.1.33

Pintrich, P. R., Smith, D. A. F., Garcia, T., & Mckeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).* http://eric.ed.gov/?id=ED338122

Pintrich, P. R., Smith, D. A. F., Garcia, T., & Mckeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, *53*(3), 801–813. https://doi.org/10.1177/0013164493053003024

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

Preston, R., Gratani, M., Owens, K., Roche, P., Zimanyi, M., & Malau-Aduli, B. (2020). Exploring the impact of assessment on medical students' learning. *Assessment & Evaluation in Higher Education*, *45*(1), 109–124. https://doi.org/10.1080/02602938.2019.1614145

Proud, S. (2015). Resits in higher education: Merely a bar to jump over, or do they give a pedagogical 'leg up'? *Assessment & Evaluation in Higher Education*, *40*(5), 681–697. https://doi.org/10.1080/02602938.2014.947241

Putwain, D. W., & Sander, P. (2016). Does the confidence of first-year undergraduate students change over time according to achievement goal profile? *Studies in Higher Education*, *41*(2), 381–398. https://doi.org/10.1080/03075079.2014.934803

Radosevich, D. J., Vaidyanathan, V. T., Yeo, S., & Radosevich, D. M. (2004). Relating goal orientation to self-regulatory processes: A longitudinal field test. *Contemporary Educational Psychology*, *29*(3), 207–229. https://doi.org/10.1016/S0361-476X(03)00032-8

Reviewcommissie Hoger Onderwijs en Onderzoek. (2016). *Eindadvies Erasmus Universiteit Rotterdam*. https://www.eur.nl/sites/corporate/files/2016-11/eindadvies_reviewcommissie.pdf

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

Ross, M. E., Salisbury-Glennon, J. D., Guarino, A., Reed, C. J., & Marshall, M. (2003). Situated self-regulation: Modeling the interrelationships among instruction, assessment, learning strategies and academic performance. *Educational Research and Evaluation*, *9*(2), 189–209. https://doi.org/10.1076/edre.9.2.189.14211

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sadler, D. R. (2007). Perils in the meticulous specification of goals and assessment criteria. *Assessment in Education: Principles, Policy & Practice*, *14*(3), 387–392. https://doi.org/10.1080/09695940701592097

Sadler, D. R. (2014). The futility of attempting to codify academic achievement standards. *Higher Education*, *67*(3), 273–288. https://doi.org/10.1007/s10734-013-9649-1

Sambell, K., & McDowell, L. (1998). The construction of the hidden curriculum: Messages and meanings in the assessment of student learning. *Assessment & Evaluation in Higher Education*, *23*(4), 391–402. https://doi.org/10.1080/0260293980230406

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In *Heijmans, R.D.H., Pollock, D.S.G., & Satorra, A. (eds.), Innovations in multivariate statistical analysis. A festschrift for Heinz Neudecker* (pp. 233–247). Kluwer Academic Publishers.

Sawyer, R. (2013). Beyond correlations: usefulness of high school GPA and test scores in making college admissions decisions. *Applied Measurement in Education*, *26*(2), 89–112. https://doi.org/10.1080/08957347.2013.765433

Schippers, M., Scheepers, A. W. A., & Peterson, J. B. (2015). *A scalable goal-setting intervention closes both the gender and ethnic minority achievement gap* (SSRN Scholarly Paper ID 2647142). Social Science Research Network. https://papers.ssrn.com/abstract=2647142

Schmidt, H. G. (1994). Problem-based learning: An introduction. *Instructional Science*, *22*(4), 247–250. https://doi.org/10.1007/BF00891778

Schmidt, H. G., Cohen-Schotanus, J., Molen, H. T. van der, Splinter, T. A. W., Bulte, J., Holdrinet, R., & Rossum, H. J. M. van. (2009). Learning more by being taught less: A "time-for-self-study" theory explaining curricular effects on graduation rate and study duration. *Higher Education*, *60*(3), 287–300. https://doi.org/10.1007/s10734-009-9300-3

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

Schunk, D. H., Meece, J. L., & Pintrich, P. R. (2014). *Motivation in Education: Theory, research and applications* (4th ed.).

Schuwirth, L., & Van Der Vleuten, C. (2004). Merging views on assessment. *Medical Education*, *38*(12), 1208–1210. https://doi.org/10.1111/j.1365-2929.2004.02055.x

Scott, E. P. (2012). Short-term gain at long-term cost? How resit policy can affect student learning. *Assessment in Education: Principles, Policy & Practice*, *19*(4), 431–449. https://doi.org/10.1080/0969594X.2012.714741

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, *35*(4), 453–472. https://doi.org/10.1023/A:1003196224280

Sebesta, A. J., & Bray Speth, E. (2017). How should I study for the exam? Self-regulated learning strategies and achievement in introductory biology. *CBE—Life Sciences Education*, *16*(2), ar30. https://doi.org/10.1187/cbe.16-09-0269

Senko, C., & Miles, K. M. (2008). Pursuing their own learning agenda: How mastery-oriented students jeopardize their class performance. *Contemporary Educational Psychology*, *33*(4), 561–583. https://doi.org/10.1016/j.cedpsych.2007.12.001

Simzar, R. M., Martinez, M., Rutherford, T., Domina, T., & Conley, A. M. (2015). Raising the stakes: How students' motivation for mathematics associates with high- and low-stakes test achievement. *Learning and Individual Differences*, *39*, 49–63. https://doi.org/10.1016/j.lindif.2015.03.002

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*(3), 417–453. https://doi.org/10.3102/00346543075003417

Sitzmann, T., & Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychological Bulletin*, *137*(3), 421–442. https://doi.org/10.1037/a0022777

Sneyers, E., & De Witte, K. (2015). The effect of an academic dismissal policy on dropout, graduation rates and student satisfaction. Evidence from the Netherlands. *Studies in Higher Education*, *42*(2), 1–36. https://doi.org/10.1080/03075079.2015.1049143

Snyder, B. R. (1971). *The hidden curriculum*. Knopf.

Stanger-Hall, K. F. (2012). Multiple-choice exams: An obstacle for higher-level thinking in introductory science classes. *CBE—Life Sciences Education*, *11*(3), 294–306. https://doi.org/10.1187/cbe.11-11-0100

Stegers-Jager, K. M., Cohen-Schotanus, J., Splinter, T. A. W., & Themmen, A. P. N. (2011). Academic dismissal policy for medical students: Effect on study progress and help-seeking behaviour. *Medical Education*, *45*(10), 987–994. https://doi.org/10.1111/j.1365-2923.2011.04004.x

Stegers-Jager, K. M., Cohen-Schotanus, J., & Themmen, A. P. N. (2012). Motivation, learning strategies, participation and medical school performance. *Medical Education*, *46*(7), 678–688. https://doi.org/10.1111/j.1365-2923.2012.04284.x

Stegers-Jager, K. M., Savas, M., Waal, J. van der, Rossum, E. F. C. van, & Woltman, A. M. (2020). Gender–specific effects of raising Year 1 standards on medical students' academic performance and stress levels. *Medical Education*, *Advance online publication*. https://doi.org/10.1111/medu.14068

Stegers-Jager, K. M., Steyerberg, E. W., Lucieer, S. M., & Themmen, A. P. N. (2015). Ethnic and social disparities in performance on medical school selection criteria. *Medical Education*, *49*(1), 124–133. https://doi.org/10.1111/medu.12536

Stegers-Jager, K. M., & Themmen, A. P. N. (2015). Binding study advice: Effect of raising the standards? *Perspectives on Medical Education*, *4*(3), 160–162. https://doi.org/10.1007/s40037-015-0180-1

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences*. Taylor & Francis.

Strikkers, H. (2015, March 5). Tweehonderd jaar studiefinanciering. *Medium*. https://medium.com/@Folia/tweehonderd-jaar-studiefinanciering-a6d883ed0748

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education*, *30*(4), 325–341. https://doi.org/10.1080/02602930500099102

Studiefinanciering door de jaren heen. (2012, October 11). *Trouw*. https://www.trouw.nl/nieuws/studiefinanciering-door-de-jaren-heen~b918398b/

Sundre, D. L., & Kitsantas, A. (2004). An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemporary Educational Psychology*, *29*(1), 6–26. https://doi.org/10.1016/S0361-476X(02)00063-2

Sungur, S. (2007). Contribution of motivational beliefs and metacognition to students' performance under consequential and nonconsequential test conditions. *Educational Research and Evaluation*, *13*(2), 127–142. https://doi.org/10.1080/13803610701234898

Tannock, S. (2017). No grades in higher education now! Revisiting the place of graded assessment in the reimagination of the public university. *Studies in Higher Education*, *42*(8), 1345–1357. https://doi.org/10.1080/03075079.2015.1092131

Taras, M. (2005). Assessment - summative and formative - some theoretical reflections. *British Journal of Educational Studies*, 466–478.

Thiede, K. W., Wiley, J., & Griffin, T. D. (2011). Test expectancy affects metacomprehension accuracy. *British Journal of Educational Psychology*, *81*(2), 264–273. https://doi.org/10.1348/135910710X510494

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education: Principles, Policy & Practice*, *14*(3). https://www.tandfonline.com/doi/full/10.1080/09695940701591867

UNESCO. (2016, January 22). *Student learning assessment and the curriculum: Issues and implications for policy, design and implementation* [Text]. International Bureau of Education. http://www.ibe.unesco.org/en/document/student-learning-assessment-and-curriculum-issues-and-implications-policy-design-and

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

Uttl, B., White, C., A., & Morin, A. (2013). The numbers tell it all: Students don't like numbers! *PLOS ONE*, *8*(12), e83443. https://doi.org/10.1371/journal.pone.0083443

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C., & Vallieres, E. F. (1992). The academic motivation scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and Psychological Measurement*, *52*(4), 1003–1017. https://doi.org/10.1177/0013164492052004025

Van den Berg, M. N., & Hofman, W. H. A. (2005). Student success in university education: A multi-measurement study of the impact of student and faculty factors on study progress. *Higher Education*, *50*(3), 413–446. https://doi.org/10.1007/s10734-004-6361-1

Van der Vleuten, C. P. M., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K. J., & Tartwijk, J. van. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher*, *34*(3), 205–214. https://doi.org/10.3109/0142159X.2012.652239

Van der Vleuten, C. P. M., Verwijnen, G. M., & Wijnen, W. H. F. W. (1996). Fifteen years of experience with progress testing in a problem-based learning curriculum. *Medical Teacher*, *18*(2), 103–109. https://doi.org/10.3109/01421599609034142

Van Herpen, S. G. A. (2019). *A head start into higher education: How students academically prepare and adjust for a successful transition into university (Doctoral dissertation)* [Erasmus University Rotterdam]. https://www.risbo.nl/_publicaties/Van%20Herpen%20(2019)%20A%20Head%20Start%20into%20Higher%20Education.pdf

Van Herpen, S. G. A., Meeuwisse, M., Hofman, W. H. A., Severiens, S. E., & Arends, L. R. (2017). Early predictors of first-year academic success at university: Pre-university effort, pre-university self-efficacy, and pre-university reasons for attending university. *Educational Research and Evaluation*, *23*(1–2), 52–72. https://doi.org/10.1080/13803611.2017.1301261

Vancouver, J. B., & Kendall, L. N. (2006). When self-efficacy negatively relates to motivation and performance in a learning context. *Journal of Applied Psychology*, *91*(5), 1146–1153. https://doi.org/10.1037/0021-9010.91.5.1146

Vermeulen, L., Scheepers, A., Adriaans, M., Arends, L., Bos, R. van den, Bouwmeester, S., Meer, F.-B. van der, Schaap, L., Smeets, G., Molen, H. van der, & Schmidt, H. (2012). Nominaal studeren in het eerste jaar. *Tijdschrift Voor Hoger Onderwijs*, *30*(3), 204–216.

Vossensteyn, H., Kottmann, A., Jongbloed, B., Kaiser, F., Cremonini, L., Stensaker, B., Hovdhaugen, E., & Wollscheid, S. (2015). *Dropout and completion in higher education in Europe: main report*. http://doc.utwente.nl/98513/

VSNU. (2012). *Prestaties in perspectief: Trendrapportage universiteiten 2000-2020*. https://www.vsnu.nl/files/documenten/Publicaties/Trendrapportage_DEF.pdf

Wolf, L. F., & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education*, *8*(3), 227–242. https://doi.org/10.1207/s15324818ame0803_3

Wood, R. E., & Locke, E. A. (1987). The relation of self-efficacy and grade goals to academic performance. *Educational and Psychological Measurement*, *47*(4), 1013–1024. https://doi.org/10.1177/0013164487474017

Woolley, K., & Fishbach, A. (2017). Immediate rewards predict adherence to long-term goals. *Personality and Social Psychology Bulletin*, *43*(2), 151–162. https://doi.org/10.1177/0146167216676480

Yocarini, I. (2019). *Testing in Higher Education: Decisions on students' performance (Doctoral dissertation)*. https://repub.eur.nl/pub/119594

Yocarini, I. E., Bouwmeester, S., Smeets, G., & Arends, L. R. (2018). Systematic comparison of decision accuracy of complex compensatory decision rules combining multiple tests in a higher education context. *Educational Measurement: Issues and Practice*. https://doi.org/10.1111/emip.12186

Zimmerman, B. J. (1986). Becoming a self-regulated learner: Which are the key subprocesses? *Contemporary Educational Psychology*, *11*(4), 307–313. https://doi.org/10.1016/0361-476X(86)90027-5

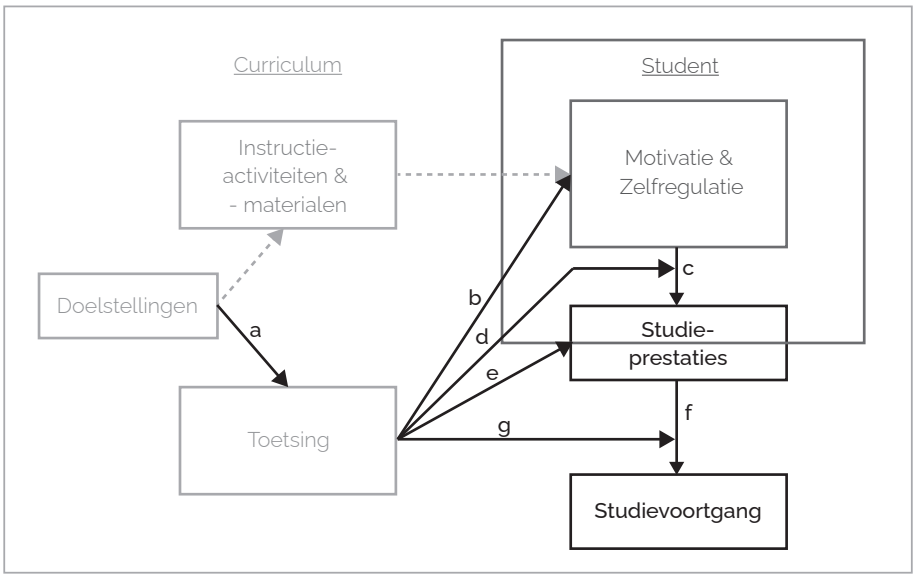
Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, *25*(1), 3–17. https://doi.org/10.1207/s15326985ep2501_2

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166–183.

SAMENVATTING

(SUMMARY IN DUTCH)

Het verbeteren van studievoortgang en studieprestaties is een voortdurende uitdaging voor opleidingen in het hoger onderwijs. Naar aanleiding hiervan was het doel van dit proefschrift om te onderzoeken of studievoortgang en studieprestaties van studenten samenhangen met kenmerken van de examensystemen in het hoger onderwijs. Aangezien motivatie en zelfregulatie twee van de belangrijkste factoren zijn die samenhangen met studieprestaties (Richardson et al., 2012), hebben we bovendien bekeken hoe de motivatie en zelfregulatie van studenten mogelijke verschillen in studieprestaties kunnen verklaren. Recente, relatief ingrijpende wijzigingen in de prestatiebelangen, prestatiestandaarden en herkansingsstandaarden van de examensystemen van alle bacheloropleidingen aan de Erasmus Universiteit Rotterdam (EUR) boden een unieke gelegenheid om de gevolgen van deze wijzigingen voor het leren van studenten (d.w.z. studievoortgang, prestaties, motivatie en zelfregulatie) te onderzoeken. In dit proefschrift zijn *prestatiebelangen* gedefinieerd als de gevolgen van onvoldoendes op tentamens, *prestatiestandaarden* beschrijven welke tentamencijfers voldoende zijn om studiepunten te behalen en *herkansingsstandaarden* betreffen het aantal toegestane herkansingen binnen een collegejaar. In deze samenvatting behandelen we de vier empirische hoofdstukken en het theoretische hoofdstuk van dit proefschrift en beschrijven we hoe onze bevindingen passen in het conceptuele model (Figuur 1) dat in hoofdstuk 1 van dit proefschrift is toegelicht.



Figuur 1. Het overkoepelende conceptuele model voor dit proefschrift. De zwarte paden a-g representeren de onderzochte associaties in de hoofdstukken 2-6: hoofdstuk 2 gaat over de paden e, f en g; hoofdstuk 3 en 4 gaan beide over de paden b, c en d; hoofdstuk 5 gaat over de paden b en c; hoofdstuk 6 gaat over de paden a, b en c. De gestippelde paden zijn weliswaar cruciaal voor het leerproces, maar niet het onderwerp van dit proefschrift.

Studievoortgang verklaren: Verschillen in prestaties en in selectie voor studievoortgang

Het belangrijkste doel voor het wijzigen van het examensysteem op de EUR was het versnellen van de studievoortgang van studenten. In de studie in **hoofdstuk 2** onderzochten we daarom verschillen in de studievoortgang onder verschillende examensystemen in het eerste studiejaar van drie grote opleidingen aan de EUR: voor bedrijfskunde (n = 2,048) was de belangrijkste aanpassing aan het examensysteem een wijziging in de prestatiebelangen; geneeskunde (n = 1,630) veranderde de prestatiebelangen en de prestatiestandaard; psychologie (n = 1,076) paste de prestatiebelangen, de prestatiestandaard en de herkansingsstandaard aan. Hoewel de wijzigingen in de prestatiestandaarden en herkansingsstandaarden verschilden tussen de opleidingen, was de verandering in prestatiebelangen in de drie opleidingen vergelijkbaar: in het oude systeem met lagere prestatiebelangen moesten studenten alle 60 studiepunten van het eerste jaar binnen twee jaar behalen, terwijl alle 60 studiepunten in het nieuwe systeem binnen één jaar moesten worden behaald.

Onze resultaten toonden aan dat studenten in alle drie de opleidingen onder het nieuwe examensysteem na één jaar een significant hogere studievoortgang hadden dan onder het oude systeem na één jaar. De studievoortgang was dus sneller onder het nieuwe systeem. De definitieve studievoortgang (d.w.z. na twee jaar onder het oude systeem versus na één jaar onder het nieuwe systeem) was onder het nieuwe systeem bij psychologie eveneens hoger, maar lager bij de opleidingen bedrijfskunde en geneeskunde. We hebben in deze studie twee verklaringen onderzocht voor de geobserveerde verschillen in studievoortgang. Verschillen in prestaties (Figuur 1, *paden e en f*) zijn een eerste mogelijke verklaring: studenten onder het nieuwe examensysteem scoorden verschillend op een aantal prestatie-indicatoren, zoals bijvoorbeeld het gemiddelde cijfer. Onze bevindingen waren over het algemeen in lijn met eerdere literatuur: hogere prestatiebelangen zijn gerelateerd aan hogere uitval, maar de prestaties van de resterende studenten zijn beter (Lindo et al., 2010). Verschillen in selectie voor studievoortgang (*pad g*) zijn een tweede mogelijke verklaring voor de studievoortgangsverschillen: een significant aantal studenten zou studievoortgang boeken onder het ene systeem, maar niet onder het andere. Het was voor de beslissing met betrekking tot de studievoortgang van studenten dus relevant welk examensysteem werd gebruikt.

Kortom, de resultaten van dit onderzoek geven aan dat de studievoortgang van studenten gevoelig is voor kenmerken van het examensysteem in verschillende opleidingen: wanneer het tijdsinterval voor het behalen van studiepunten korter is (d.w.z. hogere prestatiebelangen), tonen studenten snellere studievoortgang. Deze

verschillen in studievoortgang kunnen zowel worden verklaard door verschillen in studieprestaties, als door verschillen in selectie voor studievoortgang. Het examensysteem kan derhalve een effectieve manier zijn om de studievoortgang van studenten te stimuleren.

Prestaties verklaren: De rol van motivatie en zelf- regulatie

Ons volgende doel was om de verschillen in studieprestaties onder diverse examensystemen beter te begrijpen. In de hoofdstukken 3 en 4 hebben we daarom onderzocht hoe de motivatie en zelfregulatie van studenten hun studieprestaties konden verklaren. In de studie in **hoofdstuk 3** hebben we gebruik gemaakt van een eerder ontwikkeld en gevalideerd structureel model dat de studieprestaties van eerstejaars geneeskundestudenten probeert te verklaren (Stegers-Jager et al., 2012). In dit model is hogere motivatie gerelateerd aan betere leerstrategieën (d.w.z. zelfregulatie), hogere onderwijsparticipatie en betere studieprestaties. We vergeleken de officiële tentamencijfers en de antwoorden op een zelfrapportage-vragenlijst van studenten die ofwel studeerden onder het oude examensysteem (conjunctief, lagere prestatiebelangen, lagere prestatiestandaard; n = 648), ofwel onder het nieuwe examensysteem (compensatoir, hogere prestatiebelangen, hogere prestatiestandaard; n = 529).

De motivatie, zelfregulatie, participatie en studieprestaties van studenten waren significant hoger onder het nieuwe examensysteem. Wat betreft motivatie, rapporteerden studenten onder het nieuwe systeem hogere interesse, evenals meer geloof in eigen kunnen. Qua zelfregulatie, rapporteerden de studenten onder het nieuwe systeem over het algemeen meer diep leren (bijvoorbeeld materiaal relateren aan voorkennis, en reflecteren op het eigen leren), beter tijdsbeheer en meer doorzettingsvermogen. De gemiddelde tentamencijfers waren eveneens significant hoger onder het nieuwe systeem. Het prestatieverschil tussen studenten onder de twee examensystemen was vrijwel identiek aan het verschil in prestatiestandaarden tussen de twee systemen: een verschil van een half punt op een tien-puntschaal.

In tegenstelling tot de verschillen in gemiddelden, waren de relaties tussen motivatie, zelfregulatie, participatie en prestaties wel vergelijkbaar onder beide systemen. We concludeerden derhalve dat de betere prestaties van studenten onder het nieuwe examensysteem konden worden verklaard door hogere motivatie, zelfregulatie en participatie, maar niet door andere relaties tussen motivatie, zelfregulatie, participatie en prestaties.

In de studie in **hoofdstuk 4** hebben we wederom motivatie, zelfregulatie en studieprestaties onder twee verschillende examensystemen onderzocht, maar nu bij studenten sociale wetenschappen en gebruikmakend van de sterkste motivatie- en zelfregulatie-predictoren van studieprestaties (Richardson et al., 2012): cijferdoelen, bestaande uit zowel het streefcijfer (d.w.z. 'welk cijfer streef je na?') als het minimum cijferdoel (d.w.z. 'wat is het laagste cijfer waar je tevreden mee zou zijn?'), prestatieverwachting (d.w.z. 'welk cijfer verwacht je te krijgen? '), geloof in eigen kunnen, interesse, doorzettingsvermogen, tijdsbeheer en tentamenangst. We vergeleken twee groepen studenten die hetzelfde statistiekvak volgden onder verschillende examensystemen: psychologiestudenten (n = 219) studeerden onder een examensysteem met lagere prestatiebelangen, een lagere prestatiestandaard en minder herkansingen dan studenten pedagogiek en onderwijswetenschappen (PED; n = 85).

Wat betreft de studieprestaties, verschilde het cijfer voor het reguliere tentamen niet significant tussen studenten psychologie en PED. PED-studenten maakten echter meer gebruik van de herkansing en vertoonden hogere definitieve tentamenprestaties (na de herkansing). Het lijkt er dus op dat PED-studenten hun hogere prestaties uitstelden tot de herkansing. We vergeleken ook de motivatie en zelfregulatie die een deelsteekproef van de studenten ($n_{psychologie}$ = 150 / n_{PED} = 51) rapporteerde in een vragenlijst. Wat betreft motivatie hadden PED-studenten significant hogere minimum cijferdoelen, meer geloof in eigen kunnen en meer interesse. De streefcijfers verschilden niet significant. Aangaande zelfregulatie toonden PED-studenten een beter tijdbeheer, maar rapporteerden ze ook meer tentamenangst. Het doorzettingsvermogen verschilde daarentegen niet significant. Wat betreft de associaties van motivatie en zelfregulatie met studieprestaties vonden we geen significante verschillen. Daarnaast bleken vier van de gemeten motivatie- en zelfregulatie-variabelen een significant deel van de variantie in prestaties te verklaren: streefcijfer, prestatieverwachting, geloof in eigen kunnen, en doorzettingsvermogen.

De studies in hoofdstuk 3 en 4 vullen belangrijke lacunes in de wetenschappelijke literatuur, omdat er nauwelijks onderzoek was naar motivatie en zelfregulatie onder verschillende examensystemen. We hebben meerdere significante associaties van examensystemen met motivatie en zelfregulatie (Figuur 1, *pad b*) waargenomen bij geneeskunde- en sociale wetenschappenstudenten. Op basis van deze beide hoofdstukken concluderen we daarom dat de hogere studieprestaties onder examensystemen met hogere prestatiebelangen verklaard kunnen worden door hogere motivatie en zelfregulatie (*pad c*). Daarentegen is de relatie tussen motivatie en zelfregulatie met prestaties vergelijkbaar tussen systemen (*pad d*). Met andere

woorden, de manier waarop motivatie en zelfregulatie de prestaties beïnvloeden lijkt niet te verschillen tussen examensystemen. De belangrijkste conclusie is dat onze resultaten erop duiden dat het examensysteem kan worden gebruikt om motivatie en zelfregulatie te verhogen, zodat ook studieprestaties verbeteren.

Heeft het examensysteem op alle studenten dezelfde invloed? Latente klassen van motivatie

We vermoedden echter dat de motivatie van studenten niet altijd op dezelfde manier door het examensysteem wordt beïnvloed (d.w.z. *pad b* kan verschillen). Specifieker geformuleerd, verwachtten we dat sommige studenten meer gericht zouden zijn op het voldoen aan de minimale prestatiestandaard dan andere studenten. Daarom bekeken we in de studie in **hoofdstuk 5** hoeveel latente (d.w.z. niet direct waarneembare) klassen van studenten er bestaan in termen van cijferdoelen en prestatieverwachtingen gedurende het cruciale eerste studiejaar. We kozen cijferdoelen en prestatieverwachting om motivatie te meten, omdat dit de twee motivatiefactoren zijn die het sterkst worden geassocieerd met studieprestaties in het hoger onderwijs (Richardson et al., 2012; Schneider & Preckel, 2017). We gebruikten de blokevaluaties en de officiële tentamencijfers van studenten sociale wetenschappen (n = 587) van alle acht opeenvolgende eerstejaars blokken (d.w.z. vakken).

We vonden dat een onderverdeling in twee latente klassen de beste beschrijving van de data gaf. We noemden klasse één (66% van de studenten) de *voldoende motivatie klasse*, omdat de cijferdoelen en de prestatieverwachtingen van deze studenten het gehele jaar door dicht bij de prestatiestandaard van het examensysteem lagen. Studenten in de tweede klasse (34% van de studenten) hadden in elk van de acht eerstejaars blokken hogere cijferdoelen en prestatieverwachtingen dan de studenten uit klasse één. Daarom hebben we klasse twee de *hoge motivatie klasse* genoemd. Beide klassen vertoonden het hele jaar door relatief stabiele cijferdoelen en prestatieverwachtingen, op een daling tijdens het statistiekblok na, waar men eenmalig lagere cijferdoelen en prestatieverwachtingen rapporteerde. Het verschil tussen de twee klassen had dus niet zozeer betrekking op verschillende ontwikkelingstrajecten van cijferdoelen en prestatieverwachtingen, als wel op een verschillend niveau van motivatie gedurende het hele jaar. De validiteit van het twee-klassen model werd ondersteund door het feit dat studenten in de *hoge motivatie klasse* ook een hogere blok-interesse, blok-relevantie en blok-beoordeling rapporteerden en hogere gemiddelde cijfers behaalden dan de *voldoende motivatie klasse*. De klassen verschilden echter niet significant op gerapporteerde zelfstudietijd en ervaren hoeveelheid leren.

De studie in hoofdstuk 5 vormt een aanvulling op de eerdere studies waarin we hebben vastgesteld dat het leren van studenten sensitief bleek voor het examensysteem. De huidige studie suggereert dat de prestatiestandaard van het examensysteem vooral van belang was voor de studenten in de *voldoende motivatie klasse,* oftewel 66% van de studenten. Wanneer de prestatiestandaard zou worden verhoogd, zouden deze studenten hun motivatie moeten aanpassen, terwijl studenten uit de *hoge motivatie klasse* dit niet zouden hoeven doen. Hoewel de vraag blijft wat de percentages per klasse zouden zijn in andere studentenpopulaties, verwachten we op basis van de bevindingen in hoofdstuk 5 dat de significante verschillen in studievoortgang, studieprestaties, motivatie en zelfregulatie in de eerdere hoofdstukken voornamelijk veroorzaakt zijn door de *voldoende motivatie klasse*.

In de vier tot nu toe beschreven empirische studies hebben we vastgesteld dat studenten in examensystemen met hogere prestatiebelangen en hogere prestatiestandaarden over het algemeen gemotiveerder zijn, betere zelfregulatie vertonen, hogere cijfers halen en snellere voortgang boeken. Verder hebben we geconcludeerd dat het examensysteem vooral van belang lijkt voor de studenten in de *voldoende motivatie klasse*. Hoewel onze observatie dat het leren van studenten gevoelig lijkt voor het examensysteem niet verrassend was, waren we wel verrast door de mate van deze gevoeligheid. Voor studenten psychologie in de studie in hoofdstuk 2 was de voortgang van studenten in het nieuwe systeem bijvoorbeeld na één jaar even hoog als na twee jaar in het oude systeem. En voor de medische studenten in hoofdstuk 3 constateerden we dat de toename in prestaties de verhoging van de prestatiestandaard weerspiegelde. In beide gevallen leken de studenten zich dus volledig aan te passen aan de hogere standaarden. Naar onze mening onderstreept de omvang van de *voldoende motivatie klasse* (66% van alle studenten) de gevoeligheid van studenten voor het examensysteem.

Een verklaring voor de gevoeligheid van studenten voor het examensysteem

Om de gevoeligheid van studenten voor het examensysteem te verklaren, ontwikkelden we in de studie in **hoofdstuk 6** een theoretisch perspectief op de motivatie van studenten in het hoger onderwijs. Dit perspectief beschrijft een belangrijke reden voor de focus van studenten op tentamens, evenals de risico's van die focus. Meer specifiek stelden we dat de mate van aansluiting (*alignment*) tussen de doelstellingen van het curriculum en de tentamens een cruciale rol speelt in de motivatie van studenten. Wanneer er een perfecte aansluiting is, hebben alle doelstellingen een evenredige kans om te worden getentamineerd, en levert al het leren dus een proportionele bijdrage aan de tentamenprestaties. Aangezien leren dan een voorwaarde is voor

presteren, zou een student hetzelfde leergedrag moeten vertonen, ongeacht of de student gemotiveerd is om te leren of gemotiveerd om te presteren.

In werkelijkheid zullen tentamens echter niet perfect aansluiten op de curriculaire doelstellingen. Cohen-Schotanus (1999) legt uit dat zowel cognitieve (d.w.z. leerinhoud) als operante (d.w.z. leerhoeveelheid) aspecten van leren worden beïnvloed door tentaminering. Wij hebben dit onderscheid gebruikt om een aantal mogelijke bronnen van cognitieve en operante discrepantie (slechte aansluiting; *misalignment*) tussen doelstellingen van het curriculum en de tentamens te beschrijven. Als gevolg van die discrepantie hoeven sommige doelstellingen niet te worden behaald om te slagen voor tentamens, wat resulteert in een onderscheid tussen getentamineerde doelstellingen en niet-getentamineerde doelstellingen. Aangezien enkel de getentamineerde doelstellingen de tentamenprestaties van studenten ten goede komen, worden studenten alleen beloond voor hun inspanningen om de getentamineerde doelstellingen te behalen. Het investeren van tijd en energie in niet-getentamineerde doelstellingen wordt bovendien impliciet ontmoedigd, omdat deze investering de beschikbare tijd en energie voor het behalen van getentamineerde doelstellingen vermindert. Om de kans op goede tentamenprestaties zo hoog mogelijk te maken, is de motivatie om te presteren daarom geschikter (*more adaptive*) dan de motivatie om te leren. Met andere woorden, studenten die gemotiveerd zijn om te presteren, hebben een motivatie die past in een curriculum (*curricular fit)* waarin de tentamens niet goed aansluiten op de doelstellingen.

Deze *curricular fit* visie op motivatie heeft verschillende implicaties voor opleiders. De nadelige effecten die slecht aansluitende tentamens hebben voor leren zijn niet te onderschatten, vooral in het geval van hoge prestatiebelangen. Het is daarom zaak voor opleiders om goed te tentamineren, allereerst door de aansluiting tussen doelstellingen van het curriculum en tentamens te optimaliseren. Wanneer tentamens goed aansluiten op de doelstellingen, wordt het juiste leergedrag van studenten beloond. Bewustwording van de gevolgen die slecht aansluitende tentamens hebben voor de motivatie van studenten lijkt derhalve van belang. Bovendien kunnen opleiders kritisch beoordelen of hun tentamens gericht zijn op het weg-selecteren van de slechte presteerders, of op het stimuleren van studenten om hun volledige potentieel te ontketenen. Daarnaast hoeft het tentamen niet de eindstreep van het vak te zijn, als tentamen-inzagen een meer fundamenteel onderdeel van curricula worden. Als het onmogelijk is om tentamens goed aan te laten sluiten op curriculaire doelstellingen, zouden opleiders kunnen overwegen om niet te tentamineren, of om manieren te zoeken om de prestatiebelangen van tentamens te verminderen. Tot slot kunnen opleiders het voor studenten moeilijker maken om alleen de getentamineerde

doelstellingen te leren. Zo zou bijvoorbeeld voorafgaand aan tentamens minder informatie over de wijze van tentaminering en over de prestatiestandaard gegeven kunnen worden.

Met de beschrijving van ons theoretische model hebben we in hoofdstuk 6 een *curricular fit* perspectief op motivatie in het hoger onderwijs gepresenteerd, waarmee we de opmerkelijke gevoeligheid van studenten voor het examensysteem in de vorige hoofdstukken pogen te verklaren: in een verkeerd afgestemd curriculum is de motivatie om te presteren passender dan de motivatie om te leren, om de kans op afstuderen te maximaliseren. Vanwege de gevoeligheid van studenten voor tentamens, wordt de aansluiting van tentamens op curriculaire doelstellingen (Figuur 1, *pad a*) een noodzakelijke voorwaarde voor curricula om het leren de juiste richting op te sturen (*pad b en c*). Een belangrijk voordeel van het hanteren van een *curricular fit* perspectief op motivatie, is de resulterende focus op de invloed die onze curricula hebben op de motivatie van onze studenten.

Wat zijn de implicaties?

In **hoofdstuk 7** bespreken we een aantal implicaties van dit proefschrift. Allereerst concluderen we dat de significante verschillen in studievoortgang en studieprestaties zowel positief als negatief geïnterpreteerd kunnen worden. Aan de ene kant kan snellere studievoortgang veel tijd en energie schelen voor zowel studenten als opleiders, en kunnen betere studieprestaties betekenen dat studenten meer leren. Aan de andere kant wordt het in geval van een verhoogde focus op studievoortgang en prestaties extra belangrijk om een goede aansluiting te garanderen tussen de doelstellingen van het curriculum en de tentamens. De vraag is of deze aansluiting altijd mogelijk is en of opleiders hiervoor wel over voldoende tijd en expertise beschikken. Het in kaart brengen van deze aansluiting is daarom een cruciale richting voor toekomstig onderwijsonderzoek, maar ook zeker voor de onderwijspraktijk. Een kernvraag hierbij is: welke doelstellingen van onze curricula zijn onder-gerepresenteerd in onze tentamens? Bovenop de vraag of de aansluiting voldoende is, speelt de vraag of snelle studievoortgang wel het juiste doel is voor studenten, opleiders en de samenleving als geheel. Een hoog tempo nastreven kan immers ook schadelijke gevolgen hebben voor het leren van onze studenten, met name voor studenten uit onder-gerepresenteerde groepen in het hoger onderwijs.

Een volgende implicatie is dat het verhogen van de prestatiebelangen en prestatiestandaarden een efficiënte manier lijkt te zijn om de motivatie en zelfregulatie van studenten te verhogen. We hebben gezien dat zowel studenten geneeskunde als studenten sociale wetenschappen meer geloof in eigen kunnen, meer

interesse en beter tijdsbeheer rapporteerden onder examensystemen met hogere prestatiebelangen en -standaarden. Echter, er is ook aanleiding voor bedenkingen rondom mogelijke negatieve gevolgen van deze examensystemen. Hoe zit het met studentenwelzijn en stress? Ontnemen we studenten niet de mogelijkheid om te leren hun eigen doelen te stellen? Zijn de studenten die het niet halen wel om de juiste redenen uitgevallen? En welk niveau van motivatie is eigenlijk 'voldoende'? In ieder geval toont dit proefschrift zowel empirisch als theoretisch aan dat tentamens niet slechts het niveau van studenten bepalen nadat zij klaar zijn met leren: tentamens zijn een fundamenteel onderdeel van het curriculum met een cruciale rol bij de motivatie om te leren.

Gegeven deze cruciale rol voor motivatie en leren, bespreken we tot slot de 'juiste' prestatiebelangen, prestatiestandaard en herkansingsstandaard. Allereerst rijst de vraag wat er zou gebeuren als er nog hogere eisen gesteld zouden worden aan studenten. Wat gebeurt er bijvoorbeeld als de prestatiestandaard naar een zeven gemiddeld wordt opgehoogd? Een wellicht nog meer provocatieve optie zou zijn om de belangen en standaarden niet vooraf bekend te laten zijn, voor studenten, en mogelijk zelfs voor de opleiders. In dat geval worden de belangen en standaarden achteraf bepaald, en kunnen hierin de moeilijkheid en kwaliteit van de tentamens worden meegewogen, evenals de omstandigheden van individuele studenten. Allereerst zouden in dat geval wellicht betere beslissingen worden genomen. Bovendien kan het gevolgen voor de motivatie van studenten hebben, als er geen prestatie-ondergrens bestaat om zich op te richten.

Conclusie

De belangrijkste conclusie van dit proefschrift is dat de studievoortgang, studieprestaties, motivatie en zelfregulatie van studenten in het hoger onderwijs gevoelig zijn voor het examensysteem. Hoewel de omvang van deze gevoeligheid opmerkelijk was, is de gevoeligheid op zichzelf logisch: studenten streven ernaar om af te studeren en het examensysteem bepaalt wat daarvoor vereist is. Studenten die gevoelig zijn voor het examensysteem, vergroten daarom hun kans op afstuderen. Vanwege die gevoeligheid spelen tentamens een belangrijke rol in de motivatie van studenten en kan aanpassing van het examensysteem een zeer effectieve bron van verbeterde prestaties zijn. De mate van aansluiting tussen de curriculaire doelstellingen en tentamens bepaalt de mate waarin verbeterde prestaties een weerspiegeling zijn van verbeterd leren, zoals beoogd in de curriculaire doelstellingen. Dus, op voorwaarde dat het belang van tentamens in onze curricula tot uiting komt in de hoeveelheid tijd en energie die wordt besteed aan het garanderen van goed aansluitende tentamens, belonen cijfers het juiste leergedrag. Dan kan de gevoeligheid van studenten voor het examensysteem worden benut om het leren te verbeteren. Meer leren vergt meer inspanning, maar komt studenten op de lange termijn ten goede. Uitdagende tentamens kunnen daarom een didactische daad van liefde zijn.

CURRICULUM VITAE

Curriculum Vitae

Robert Kickert was born in Heemstede, The Netherlands, on May 23rd 1985. After completing his secondary (pre-university; VWO) education at Philips van Horne in Weert in 2003, he studied Psychology at Erasmus University Rotterdam (EUR). He obtained his Master's diploma in Biological and Cognitive Psychology in 2008. While finishing his Master's, Rob already started his teaching career, as a tutor in Psychology in 2007. After one year, he took a year off to travel the world, after which he returned as a tutor for two more years. In the summer of 2011, Rob joined the team of Pedagogical and Educational Sciences, a course programme that would start for the first time later that same year. There, Rob developed his passion for teaching and education by performing a wide range of educational activities. He worked as a student advisor, tutor, tutor trainer, tutor coach, teacher-advisor on problem-based learning, coordinator of skills trainings (in problem-based learning, presenting, SPSS, debating, and qualitative research), as well as a coordinator of the first-year statistics course. In April 2015, Rob started his PhD in Educational Sciences. During this PhD, he was a tutor in several statistics courses and one course on biological psychology, as well as tutor coordinator in the Master's statistics course and course coordinator in the first-year statistics course. In addition, he was a member of the educational committee of the Interuniversity Center for Educational research (ICO) for more than 3 years, and the chair of the educational committee as well as an ICO board member for two years. He presented his research at several (inter)national conferences, served as a reviewer, and followed multiple courses of both ICO and the Erasmus Graduate School of the Social Sciences and the Humanities. During his time at EUR, Rob found various ways to express his organisational engagement and enthousiasm: he organised three team outings, two Graduate Research Days, as well as numerous PhD Ping-Pong evenings and team drinks. He also received one award: The Graduate School Award for PhD Excellence: PhD-Colleague of the Year 2018.

Publications

Kickert, R., Meeuwisse, M., Stegers-Jager, K. M., Koppenol-Gonzalez, G. V., Arends, L. R., & Prinzie, P. (2019). Assessment policies and academic performance within a single course: The role of motivation and self-regulation. *Assessment & Evaluation in Higher Education*, *44*(8), 1177–1190. https://doi.org/10.1080/02602938.2019.1580674

Kickert, R., Stegers-Jager, K. M., Meeuwisse, M., Prinzie, P., & Arends, L. R. (2018). The role of the assessment policy in the relation between learning and performance. *Medical Education*, *52*(3), 324–335. https://doi.org/10.1111/medu.13487

Papers

Kickert, R., Meeuwisse, M., Arends, L.R., Prinzie, P., & Stegers-Jager, K.M (submitted). Assessment policies and academic progress: Differences in performance and selection for progress.

Kickert, R., Meeuwisse, M., Jongerling, J., Stegers-Jager, K.M, Arends, L.R., & Prinzie, P. (submitted). Grade goals and performance self-efficacy throughout the first academic year: A latent class approach.

Kickert, R., Meeuwisse, M., Stegers-Jager, K.M., Prinzie, P., & Arends, L.R. (submitted). Curricular fit perspective on motivation in higher education.

Yocarini, I.E., **Kickert, R.,** Jongerling, J, Meeuwisse, M., & Bouwmeester, S. (in preparation). Trackin higher education students' study time investment intensively.

Faber, T.J.E, Dankbaar, M.E.W, Van den Broek, W.W., **Kickert, R.,** & Van Merrienboer, J.J.G.M. (in preparation). Self-regulated learning and in-game behaviour as indicator for game-based learning in emergency care.

Presentations

Kickert, R., Meeuwisse, M., Prinzie, P., Arends, L.R. & Stegers-Jager, K.M. (2015, June 19). *De invloed van ' Nominaal is Normaal' (N=N) op de tijdbesteding van studenten.* Leiden, Onderwijs Research Dagen.

Kickert, R., Stegers-Jager, K.M., Meeuwisse, M., Prinzie, P. & Arends, L.R. (2016, May 25). *De rol van motivatie, leerstrategieën en participatie in het behalen van studieprestaties onder een vernieuwd examensysteem.* Rotterdam, Onderwijs Research Dagen.

Kickert, R., Stegers-Jager, K.M., Meeuwisse, M., Prinzie, P. & Arends, L.R. (2016, July 13). *The role of motivation, learning strategies and participation in academic performance under a renewed examination system: refitting a structural model.* Amsterdam, Higher Education Conference.

Kickert, R., Meeuwisse, M., Stegers-Jager, K.M., Prinzie, P. & Arends, L.R. (2016, November 2). *Academic progress under various assessment regimes.* Bad Schussenried, ICO International Fall School 2016.

Kickert, R., Stegers-Jager, K.M., Meeuwisse, M., Arends, L.R. & Prinzie, P. (2017, February 21). *De rol van zelfregulerend leren en participatie in het behalen van studieprestaties onder een vernieuwd toetsingssysteem.* Rotterdam, Erasmus Education Research (EER) lunchmeeting.

Kickert, R., Stegers-Jager, K.M., Meeuwisse, M., Arends, L.R. & Prinzie, P. (2017, March 30). *De rol van zelfregulerend leren en participatie in het behalen van studieprestaties onder een vernieuwd toetsingssysteem.* Breda, Avans Hogeschool, Kenniskring.

Kickert, R., Stegers-Jager, K.M., Meeuwisse, M., Arends, L.R. & Prinzie, P. (2017, April 20). *The role of the assessment system in the relation between learning and performance.* Utrecht, ICO National Spring School.

Kickert, R., Stegers-Jager, K.M., Meeuwisse, M., Arends, L.R. & Prinzie, P. (2017, April 25). *The role of the assessment system in the relation between learning and performance.* Rotterdam, Erasmus Medical Center, IMERR labmeeting.

Kickert, R., Stegers-Jager, K.M., Meeuwisse, M., Arends, L.R. & Prinzie, P. (2017, June 9). *The role of the assessment system in the relation between learning and performance.* Rotterdam, Erasmus University College, CEL innovation room.

Kickert, R., Meeuwisse, M., Stegers-Jager, K.M., Prinzie, P. & Arends, L.R. (2017, August 23). *Academic progress under various assessment systems.* Kopenhagen, ECER conference.

Kickert, R. (2018). The role of the assessment policy in the relation between learning and performance - Rob Kickert Interview. Medical education podcasts.

Kickert, R., Stegers-Jager, K.M., Meeuwisse, M. & Arends, L.R. (2018, april 15). *The role of the assessment policy in the relation between learning and performance.* New York, AERA conference.

Kickert, R., Meeuwisse, M., Stegers-Jager, K.M., Prinzie, P. & Arends, L.R. (2019, March 14). *Motivation: Curricular fit.* Rotterdam, DPECS Graduate Research Day

Kickert, R., Meeuwisse, M., Stegers-Jager, K.M., Prinzie, P. & Arends, L.R. (2019, May 14). *Motivation, self-regulation and academic performance under various assessment policies.* Rotterdam, CLI research lunch.

Kickert, R., Meeuwisse, M., Stegers-Jager, K.M., Prinzie, P. & Arends, L.R. (2019, August 16). *Assessment policies and academic performance: The role of motivation and self-regulation.* Aachen, EARLI Conference.

Kickert, R., Meeuwisse, M., Stegers-Jager, K.M., Prinzie, P. & Arends, L.R. (2019, August 15). *The curricular fit theory of motivation.* Aachen, EARLI Conference.

Courses

ICO Introductory Course (Spring 2016)	5 ECTS
ICO International Fall School (2016)	3 ECTS
ICO National Spring School (2017)	1 ECTS
EGSH English Academic Writing Course (2017)	2.5 ECTS
ICO Course Research into Higher Education (2017)	3 ECTS
ICO Course Assessment, Evaluation and Examination (2017)	3 ECTS
EGSH Great Thinkers of the 20th Centrury Course (2018)	2.5 ECTS
ICO Course Qualitative Research, Advanced (2019)	3 ECTS
EGSH Professionalism and Integrity in Research Course (2019)	1 ECTS

ICO dissertation series

403. Wijnen, M. (01-02-2019*) Introduction of problem-based learning at the Erasmus School of Law: Influences on study processes and outcomes*. Rotterdam: Erasmus University Rotterdam
404. Dobbelaer, M.J. (22-02-2019) *The quality and qualities of classroom observation systems*. Enschede: University of Twente
405. Van der Meulen, A.N. (28-02-2019) *Social cognition of children and young adults in context*. Amsterdam: Vrije Universiteit Amsterdam
406. Schep, M. (06-03-2019) *Guidance for guiding. Professionalization of guides in museums of art and history.* Amsterdam: University of Amsterdam
407. Jonker, H.M. (09-04-2019) *Teachers' perceptions of the collaborative design and implementation of flexibility in a blended curriculum*. Amsterdam: University of Amsterdam
408. Wanders, F. H. K. (03-05-2019). *The contribution of schools to societal participation of young adults: The role of teachers, parents, and friends in stimulating societal interest and societal involvement during adolescence.* Amsterdam: University of Amsterdam
409. Schrijvers, M.S.T. (03-05-2019) *The story, the self, the other. Developing insight into human nature in the literature classroom*. Amsterdam: University of Amsterdam
410. Degrande, T. (08-05-2019) *To add or to multiply? An investigation of children's preference for additive or multiplicative relations*. Leuven: KU Leuven.
411. Filius, R.M. (23-05-2019) *Peer feedback to promote deep learning in online education. Unravelling the process.* Utrecht: Utrecht University
412. Woldman, N. (24-05-2019) *Competence development of temporary agency workers*. Wageningen: Wageningen University
413. Donszelman, S. (06-06-2019) *Doeltaal-leertaal didactiek, professionalisering en leereffecten*. Amsterdam: Vrije Universiteit Amsterdam
414. Van Oeveren, C.D.P. (12-06-2019) *ITHAKA gaf je de reis*. Amsterdam: Vrije Universiteit Amsterdam
415. Agricola, B.T. (21-06-2019) *Who's in control? Finding balance in student-teacher interactions*. Utrecht: Utrecht University
416. Cuyvers, K. (28-08-2019), *Unravelling medical specialists self -regulated learning in the clinical environment*. Antwerp: University of Antwerp
417. Vossen, T.E. (04-09-2019*) Research and design in STEM education.* Leiden: Leiden University
418. Van Kampen, E. (05-09-2019) *What's CLIL about bilingual education*? Leiden: Leiden University

419. Henderikx, M.A. (06-09-2019) *Mind the Gap: Unravelling learner success and behaviour in Massive Open Online Courses*. Heerlen: Open University of the Netherlands
420. Liu, M. (13-09-2019) *Exploring culture-related values in Chinese student teachers' professional self-understanding and teaching experiences*. Utrecht: Utrecht University
421. Sun, X. (13-09-2019) *Teacher-student interpersonal relationships in Chinese secondary education classrooms*. Utrecht: Utrecht University
422. Wu, Q. (02-10-2019) *Making Construct-Irrelevant Variance Relevant: Modelling item position effects and response behaviors on multiple-choice tests*. Leuven: KU Leuven
423. Jansen, R.S. (11-10-2019) *Dealing with autonomy: Self-regulated learning in open online education*. Utrecht: Utrecht University
424. Van Ginkel, S.O. (23-10-2019) *Fostering oral presentation competence in higher education*. Wageningen: Wageningen University
425. Van der Zanden, P. (05-11-2019) *First-year student success at university: Domains, predictors, and preperation in secondary education*. Nijmegen: Radboud University Nijmegen
426. De Bruijn, A.G.M. (14-11-2019) *The brain in motion: Effects of different types of physical activity on primary school children's academic achievement and brain activation*. Groningen: University of Groningen
427. Hopster-Den Otter, D. (28-11-2019) *Formative assessment design: A balancing act*. Enschede: University of Twente
428. Harmsen, R. (10-12-2019) *Let's talk about stress. Beginning secondary school teachers' stress in the context of induction programmes*. Groningen: University of Groningen
429. Post, T. (11-12-2019) *Fostering inquiry-based pedagogy in primary school: a longitudinal study into the effects of a two-year school improvement project*. Enschede: University of Twente
430. Ackermans, K. (20-12-2019) *Designing Video-Enhanced Rubrics to Master Complex Skills*. Heerlen: Open University of the Netherlands

DANKWOORD

(ACKNOWLEDGEMENTS IN DUTCH)

Vijf jaar blijkt bij nader inzien een vrij aanzienlijke periode om aan onderzoek naar één onderwerp te werken; afgerond is het immers een decennium. Derhalve is er een hele rits aan mensen zonder wie ik deze periode niet (zo prettig)[1] was doorgekomen. Dit dankwoord is mijn per definitie ontoereikende poging om de dankbaarheid voor al jullie steun in woorden om te zetten[2].

Aangezien de inleiding van mijn proefschrift een ode was aan mijn vader[3], kan dit dankwoord maar op één manier beginnen: met mijn moeder. Ma, ik heb niet alleen een aanzienlijk (doch ongespecificeerd) deel van mijn opgewekte aard en persoonlijkheid aan jou te danken, maar ook zeer zeker vrijwel alle mooie interacties die ik met andere mensen heb, en dat zijn er nogal wat. Jij gaf me middels jouw grenzeloze liefde een basale vorm van eigenwaarde en zelfvertrouwen waar ik dagelijks op kan bouwen. Daarmee ben je ook gelijk de belangrijkste reden dat ik pa's 'eigenzinnige opvoedstijl' heb overleefd… Pa, jij hebt het ook 'aardig'[4] gedaan[5]. Ik zou iedereen ouders zoals jullie gunnen die ze liefdevol helpen en bewonderend aanmoedigen om zich door de roerigheden des leven heen te manoeuvreren. En dan heb ik met Paulus en De Bron ook nog eens twee 'grote' broers die van jongs af aan in alle opzichten over me gewaakt hebben. Bovendien zijn we later versterkt met nog twee geweldenaars, Lisa en Alex. Wat ben ik blij dat ik een deel ben van ons gezin.

Ik kan hetzelfde zeggen over al mijn andere vrienden en familie: mijn grootste kracht is het netwerk waar ik onderdeel van uitmaak. Wat ben ik een geluksvogelt[6]! Dat de meesten van jullie geen kaas hebben gegeten van geavanceerde statistische analyses zie ik derhalve volgaarne door de vingers. Ik heb hier niet de ruimte om alle liefde die ik van jullie ontvang de eer te geven die het verdient, dus doe ik dat (met jullie goedkeuring[7]) maar 'gewoon' middels elke interactie die we nog met elkaar zullen hebben[8]. Het voelt bijvoorbeeld vreemd om Mano, Nicolini, Tobin en Celia niet te

1 Het gebruik van haakjes en voetnoten is mij ontraden v.w.b. de wetenschappelijke onderdelen van dit proefschrift. De oplettende lezer zal merken dat ik een inventieve manier heb gevonden om mijn haakjes- en voetnotendrang toch te kunnen botvieren op u, de lezer.

2 Mocht iemand zich onvoldoende bedankt voelen middels dit dankwoord, dan hoop ik dat diegene zijn of haar percepties van onze overige interacties zal laten prevaleren boven de tekortschietende woorden van dank alhier.

3 Ik besef mij terdege dat niet elke lezer dit als zodanig geïnterpreteerd heeft.

4 ;-)

5 En volgens mij kan ik mijn dank niet beter uitdrukken dan door de inleiding van mijn proefschrift aan jou te wijden, dat zegt alles over wat je voor mij betekent.

6 Uit Rotterdam dus.

7 Die bij deze, stilzwijgend, is verleend.

8 Interacties waaruit dit niet blijkt zullen derhalve getypeerd worden als: de uitzonderingen die de regel bevestigen.

bedanken voor alle maandagavonden dat ik deel uit mocht maken van jullie gezin, of al mijn hyena-vrienden voor de mooie vakanties en ontspanning, of Jelle voor het verzorgen van een epische slotfase, of Hanna voor je buitengewone[9] nazorg[10]; dit dankwoord wordt dan simpelweg te langdradig. Mijn enorme waardering voor de vriendschap van mooie mannen als Tom, Koen, Patrick, Michaël, Teun, Adrie, Moker, Daan, Maarten en Aljosja kan ik daarom ook niet expliciet vermelden hier. Jammer!

Voor het planten van het promotie-zaadje moet ik allereerst Fabie bedanken, daar jij mij overtuigend aanzette tot het beginnen van een promotietraject; iets waar ik tot dan toe niet aan moest denken[11]. Lyanda, bedankt dat je dit aanvankelijke idee wist te laten ontsporen richting de onderwijswetenschappen en voor jouw drang om mij te behouden voor de opleiding. Estella, bedankt voor je kritische vragen en je steun bij het aanvangen van dit project.

Toen brak het moment aan waar mijn promoteam[12] ten tonele verscheen, daar jullie bereid waren mij onder jullie hoede te nemen. Ik snap dankzij jullie dat autonomie een zorgvuldige combinatie is van erg veel steun, een positief kritische houding en erg veel vrijheid om eigen keuzes te maken. Ik besef me terdege dat niet iedere PhD hiermee gezegend is; mijn dank is groot. Daarnaast wil ik jullie uit de grond van mijn hart bedanken voor de talloze keren dat ik met afgrijzen heb aanschouwd hoe ik mijn (in eigen ogen sublieme) schrijfwerk retour ontving als een kakafonie van gekleurde ballonnetjes, waar ik me binnensmonds vloekend een weg door moest banen. Jullie inhoudelijke scherpte kende geen grenzen, zowel op details als hoofdlijnen: elke zwakke plek werd door jullie geïdentificeerd. Dit was zeker niet altijd een onmiddellijk genoegen, maar ik waardeer het desalniettemin zeer; de mensen die je opbouwende kritiek geven zijn immers de mensen die je verder helpen. Marieke, je hebt als een wetenschapsmoeder over me gewaakt deze jaren. Je stond voor me klaar in emotionele, inhoudelijke en praktische zin, met een luisterend oor, oprechte interesse, hart voor de zaak en een enorme gevoeligheid. Bedankt voor al je tijd en steun, je bent wat mij betreft een lichtend voorbeeld voor alle dagelijks begeleiders. Karen, ook vanuit jou voelde ik altijd enorme lieve steun, evenals waardering voor wie ik ben en wat ik kan. Je was vaak degene die met mooie, terechte, kritische discussiepunten op de proppen kwam, en daar draait de wetenschap wat mij betreft om. Lidia, bedankt voor je opgewekte enthousiasme, je scherpe ondervragingen over mijn 'zweverige' ideeën, je geloof in mijn docentvaardigheden en je altijd humoristische inslag. Peter,

9 Zo buitengewoon dat je eigenlijk in elk dankwoord wel even genoemd zou moeten worden in mijn ogen.
10 Die hopelijk nog heel erg lang mag duren.
11 Wijsheid komt blijkbaar niet altijd met de jaren… ;-)
12 Uit te spreken met een natte t.

bedankt voor het ondersteunen van mijn competentie en autonomie, en voor onze verbondenheid. Dit[13] begon al voor de start van dit promotietraject en hield jij vol tot aan het einde: je bent er echt voor me geweest. Ik wil jullie alle vier enorm bedanken dat ik me altijd veilig heb gevoeld om fouten te maken[14], en vooral ook mezelf te zijn.

Ik wil Cees van der Vleuten, Janke Cohen-Schotanus en Guus Smeets heel erg bedanken voor de tijd en moeite om mijn proefschrift te lezen en (positief) te beoordelen. Guus wil ik daarnaast bedanken voor je geloof in mijn kunnen, en je explicitering daarvan. Ik wil iedereen bij ICO, en Caroline in het bijzonder, bedanken voor alle leerzame cursussen en voor (het bieden van een podium voor mijn drang om me met onderwijs te bemoeien, middels) de bijeenkomsten met de onderwijscommissie en bestuursraad.

Joran, hoewel jij officieel geen begeleider was voelde het soms wel zo. Bedankt voor je vriendschap, voor de koffie en grappen, voor de lichtzinnige doch kritische reflecties op het universitaire reilen en zeilen en voor de talloze keren dat je wat uitmuntend statistisch advies uit je mouw schudde, alsof het niets was. I respect ya brah. Dat geldt ook voor de rest van de statistieksectie: bedankt voor de fijne overleggen vol humor. In het bijzonder wil ik 'Peter II'[15] bedanken voor de ontspannen samenwerking tijdens 1.3. En Gaab, bedankt voor je vriendschap, voor het delen van onze zelfreflecties en voor je enthousiaste analyse-hulp.

Van mijn minder directe collegae wil ik allereerst 'Peter III'[16] bedanken voor je oeverloze hulp met het opschonen van mijn datasets. Dit waren de meest verwarrende en verontrustende momenten van mijn werkende tijd bij de EUR, waar ik dankzij jou met een glimlach aan terugdenk. Sanne, bedankt voor onze fijne koffiegesprekken, je advies en je prettige neer-op-aardse[17] houding. Sanneke, bedankt voor de gezelligheid op congres en T16-12. Eva en Loes, bedankt voor de extracurriculaire gezelligheid. Ik vreesde op een gegeven moment dat ik geen episch feestcongres mee zou gaan maken, maar daar hebben jullie mede een stokje voor gestoken in NYC.

13 Het is inderdaad onduidelijk waar 'dit' naar verwijst ja :-).

14 Hoewel daar tot op heden geen geregistreerd geval van is, heb ik de theoretische mogelijkheid erg gewaardeerd… ;-)

15 Deze telling is binnen een robocentrisch wereldbeeld tot stand gekomen, ik zou willen aanmoedigen dat je jezelf als de eerste Peter blijft beschouwen.

16 Zie noot 15.

17 Vrij vertaald.

Rianne, Sabine, Jacqueline, Maartje, Rian, Lonneke, Roel, Emily, Adriaan, Gino, Frank, Wendy Tie, Kayin, Hester, Wendy Te en alle anderen van de opleiding pedagogiek en onderwijswetenschappen, jullie hebben de EUR mijn thuis gemaakt. Brian, bedankt voor je flauwe grappen[18] en onze kritische inhoudelijke gesprekken; Iliass, bedankt voor het vertrouwen en de kansen die je me bood. Mireille, bedankt voor al het dissen en dat je zo lief bent, hoe lelijk je truien ook mogen zijn.

En zo kom ik aan bij mijn lotgenoten, de PhDs van T16. Ik ben elk van jullie dankbaar voor de sfeer op T16, de goede gesprekken en de fenomenale grappen, maar wil er een aantal even uitlichten. First of all, I immensely appreciated our pingpong extravaganzas[19]. At many times, these kept me going. So thanks to all of you (for all your staggering defeats), particularly several members of the hardcore crew: Marcelo, Peiqian, Wei, Job, André, Omer, Sergio, and Maike. Anniek, Aike, Donna, Ilse, Jacqueline, Andrea, Gloria, Denise, Lara, Zhiqi, Maria, Fleur and Nouran, although your pingpong-enthousiasm could use a boost (to various degrees), I greatly appreciate our personal talks, interesting discussions and loud[20] laughs; Julia, daar bovenop nog bedankt dat je zo'n ontzettende lieve knuffelbeer bent. Willemijn en Sabrina, niet alleen bedankt voor de pingpong en alle lol, maar ook voor de goede gesprekken en bakken emotionele rugdekking door de jaren heen. Iris, bedankt dat je een bruisend toonbeeld van energie en ondernemendheid was[21]. Lara, bedankt voor je subtiele schoppen onder mijn kont, voor je stijltips, voor de wereldwijde gezelligheid en vooral voor het feit dat ik nu die blik van jou voor me zie die je trekt als je me net een snoeiharde burn hebt gegeven.

Ik wil mijn roomies bedanken dat jullie het met me hebben uitgehouden, zonder jullie gezelligheid had ik het niet eens willen redden. Arend, jij bent het levende bewijs dat niet elke goede wetenschapper gepromoveerd hoeft te zijn[22]. Tessa en Robert, bedankt voor de korte, maar gezellige afsluitingstijd. Eke, bedankt voor de felle discussies en je enorme drive om van de wereld een eerlijker plek te maken. Ildeniz, ik bewonder je onuitputtelijke vrolijkheid en je lieve sarcasme[23].

D

18 Tenminste, ik neem aan dat je niet serieus voor Feyenoord bent toch?!

19 I know, it's hard to tell when I enjoy something…

20 Or maybe that was just me…?

21 Oftewel 'de yocarini' van de afdeling.

22 Ik ontken met kracht elke vorm van betrokkenheid bij jouw keuze om te stoppen ;-).

23 Jij kan het zo lief en achteloos brengen dat het soms een paar minuten duurde voor ik doorhad wat je écht zei.

Tot slot: wie had gedacht dat een promotie ook tot gezinsuitbreiding kan leiden?! Ik kreeg er immers twee broers bij. Broers, als er twee mensen zijn die me hier doorheen gesleept hebben dan zijn jullie het. Met alle gesynchroniseerde pauzes, ad hoc overlegjes, met lach doorspekte lunches, gepassioneerde pingpong, zielengeroer, kont-schoppen en bol-aaitjes, kritische kanttekeningen, Lidl fantalk, hartverwarmende lachsalvo's, fenomenale non-verbale gesprekken en overige briljante grappen hebben jullie jezelf zowel Kroeloematig als Conceptueel overstegen[24]. Ik kijk tegen jullie op. En vind het sierend dat jullie deze woorden waarschijnlijk licht gegeneerd[25] gaan ontvangen.

Om een lang verhaal kort te maken:

Mijn leven is een dans op de schouders van reuzen.

24 Dit zeg ik echt niet alleen maar omdat jullie je dankwoord nog moeten gaan schrijven…

25 Ik denk aan: met een blos op de wangen en de blik neerwaarts, richting de naar elkaar toegedraaide voeten.