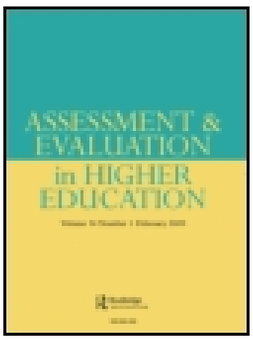

Routledge
Taylor & Francis Group

Assessment & Evaluation in Higher Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/caeh20

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

Rob Kickert , Marieke Meeuwisse , Lidia R. Arends , Peter Prinzie & Karen M. Stegers-Jager

To cite this article: Rob Kickert , Marieke Meeuwisse , Lidia R. Arends , Peter Prinzie & Karen M. Stegers-Jager (2020): Assessment policies and academic progress: differences in performance and selection for progress, Assessment & Evaluation in Higher Education, DOI: 10.1080/02602938.2020.1845607

To link to this article: https://doi.org/10.1080/02602938.2020.1845607

Published online: 23 Nov 2020.

ASSESSMENT & EVALUATION IN HIGHER EDUCATION
https://doi.org/10.1080/02602938.2020.1845607

Routledge
Taylor & Francis Group

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

Rob Kickert[a], Marieke Meeuwisse[a], Lidia R. Arends[a,b], Peter Prinzie[a] and Karen M. Stegers-Jager[c]

[a]Department of Psychology, Education & Child Studies, Erasmus School of Social and Behavioural Sciences, Erasmus University Rotterdam, Rotterdam, The Netherlands; [b]Department of Biostatistics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands; [c]Institute of Medical Education Research Rotterdam, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

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, 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

Introduction

Swift academic progress saves time, money and energy for students, educators and society. Therefore, optimising academic progress is an important goal for educational stakeholders worldwide (Attewell, Heil, and Reisel 2011; Vossensteyn et al. 2015). Adapting characteristics of assessment policies may be an effective way to improve academic progress, given the premises that (i) characteristics of assessment policies are related with student grades (Cole and Osterlind 2008; Elikai and 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'

CONTACT Rob Kickert r.kickert@essb.eur.nl

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 (Wolf and Smith 1995; Cole and Osterlind 2008). Consequently, raising the stakes might be an efficacious way to enhance academic progress. Research on the use of academic probation for low-performing students shows that probation encourages some students to drop out, while improving grades for those students who decide to stay in the course programme (Lindo, Sanders, and Oreopoulos 2010). On the other hand, previous research on Dutch assessment policies showed higher first-year dropout rates (Arnold 2015; Sneyers and De Witte 2017), as well as lower grades (De Koning et al. 2014) under academic dismissal policies. Results on academic progress were mixed, showing either no increase in progress (Stegers-Jager et al. 2011; Eijsvogels et al. 2015), 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 the case of conjunctive performance standards, students need to pass each individual assessment with a satisfactory grade (Chester 2005).

On the one hand, higher performance standards have consistently been associated with higher grades (Johnson and Beck 1988; Elikai and Schuhmann 2010; 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 and 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 standards

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 and 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 (Pell, Boursicot, and Roberts 2009; 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 and 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

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). For instance, for students who studied under a lower-stakes assessment policy with a conjunctive performance standard (e.g. students need a 5.5 on a 10-point scale for each individual assessment to progress), we mimicked progress under the compensatory performance standard of the higher-stakes policy (e.g. students need a 6.0 average to progress). As a result, we could now see whether these students progressed under the lower-stakes policy performance standard, as well as under the higher-stakes policy performance standard. 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 from three course programmes that changed their assessment policies in order to accelerate academic progress: business administration, medicine and psychology. We chose these three course programmes because of the large numbers of enrolling students, as well as a lack of other changes to their curricula, as was the case in other programmes of this university. 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, and 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 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 (although in medicine, only students with less than 40 credits who failed to attend compulsory support meetings were dismissed), 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. Within the psychology curriculum, a distinction exists between knowledge courses and skills training. In the lower-stakes psychology policy the performance standard and resit standard were different for the skills and knowledge assessments.

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
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.

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, Scheepers, and Peterson 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).

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 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 one 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 and 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 versus all higher-stakes policy students, and a t-test comparing only the students who progressed under the lower-stakes versus the higher-stakes policy. We calculated Cohen's *d* as a measure of effect size (0.20/0.50/0.80 = 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). 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 versus the higher-stakes policy for each course programme (RQ1). For *business administration*, one-year progress in the higher-stakes assessment policy (52.4%) was significantly higher than in the lower-stakes policy (31.4%), $\chi^2(1) = 79.01$, $p < .001$, $OR_{progress\ lower\text{-}stakes/higher\text{-}stakes} = 0.41$. Final progress

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) $N = 729$

in the higher-stakes policy (52.4%) was significantly lower than final progress in the lower-stakes policy (64.0%), $\chi^2(1) = 24.59$, $p < .001$, $OR_{progress\ lower\text{-}stakes/higher\text{-}stakes} = 1.62$ (see Table 2).

For *medicine*, students in the higher-stakes assessment policy showed significantly higher one-year progress (71.1%) than students in the lower-stakes policy (50.9%), $\chi^2(1) = 70.00$, $p < .001$, $OR_{progress\ lower\text{-}stakes/higher\text{-}stakes} = 0.42$. However, final progress was significantly lower in the higher-stakes policy (71.1%) than in the lower-stakes policy (85.5%), $\chi^2(1) = 49.73$, $p < .001$, $OR_{progress\ lower\text{-}stakes/higher\text{-}stakes} = 2.40$ (see Table 3).

Psychology students' one-year progress in the higher-stakes assessment policy (74.7%) was significantly higher than one-year progress in the lower-stakes policy (51.6%), $\chi^2(1) = 61.30$, $p < .001$, $OR_{progress\ lower\text{-}stakes/higher\text{-}stakes} = 0.36$. Final progress was also significantly higher in the higher-stakes policy (74.7%) than in the lower-stakes policy (68.8%), $\chi^2(1) = 4.59$, $p = .032$, $OR_{progress\ lower\text{-}stakes/higher\text{-}stakes} = 0.75$ (see Table 4).

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 ($M = 6.52$) had a significantly higher GPA than higher-stakes policy students ($M = 6.41$), $t(1480.35) = 2.17$, $p = .030$, $d = .10$. After selecting only the (final) progressing students, lower-stakes policy students ($M = 7.07$) showed a significantly lower GPA than higher-stakes policy students ($M = 7.15$), $t(940.12) = -2.45$, $p = .014$, $d = -.15$ (Table 2).

For *medicine*, we did not find a statistically significant difference between the GPA of all lower-stakes policy students ($M = 6.38$) and all higher-stakes policy students ($M = 6.31$), $t(1551.38) = 1.46$, $p = .143$, $d = .07$. When comparing the GPA of progressing students, lower-stakes policy students ($M = 6.62$) showed a significantly lower GPA than higher-stakes policy students ($M = 6.82$), $t(1159.16) = -5.92$, $p < .001$, $d = -.34$ (Table 3).

For *psychology*, when comparing all students, the knowledge GPA was significantly lower under the lower-stakes policy ($M = 5.92$) than under the higher-stakes policy ($M = 6.37$), $t(1067.30) = -6.20$, $p < .001$, $d = -0.38$. However, the skills GPA was significantly higher for lower-stakes policy students ($M = 7.20$) than for higher-stakes policy students ($M = 6.91$), $t(868.30) = 6.60$, $p < .001$, $d = 0.41$. After selecting the progressing students, lower-stakes policy students ($M = 6.42$) still showed a significantly lower knowledge GPA than higher-stakes policy students ($M = 6.83$), $t(701.45) = -7.21$, $p < .001$, $d = -0.52$. Again, the skills GPA was significantly higher for lower-stakes policy students ($M = 7.37$) than for higher-stakes policy students ($M = 7.21$), $t(746.69) = 5.61$, $p < .001$, $d = 0.40$ (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

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.	
									No	Yes
							Progress HSP P.S.	No	110	47
								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.	
									No	Yes
							Progress HSP P.S.	No	237	1
								Yes	187	396

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

under the higher-stakes policy. For medicine, under the performance standards of the lower-stakes assessment policy, lower-stakes policy students (85.7%) showed significantly higher progress than higher-stakes policy students (48.4%), $\chi^2(1) = 255.98$, $p < .001$, $OR_{progress\ lower\text{-}stakes/higher\text{-}stakes} = 6.38$. Under the performance standards of the higher-stakes assessment policy, lower-stakes policy students (80.6%) also showed significantly higher progress than higher-stakes policy students (71.0%), $\chi^2(1) = 20.38$, $p < .001$, $OR_{progress\ lower\text{-}stakes/higher\text{-}stakes} = 1.70$. 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).

For psychology, under the performance standards of the lower-stakes assessment policy, lower-stakes policy students' progress (34.2%) did not differ significantly from higher-stakes policy students' progress (36.7%), $\chi^2(1) = 0.71$, $p = .401$, $OR_{progress\ lower\text{-}stakes/higher\text{-}stakes} = 0.90$. Under the performance standards of the higher-stakes assessment policy, lower-stakes policy students (45.9%) showed significantly lower progress than higher-stakes policy students (74.9%), $\chi^2(1) = 94.18$, $p < .001$, $OR_{progress\ lower\text{-}stakes/higher\text{-}stakes} = 0.28$. 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).

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	
											No	Yes
									HSP P.S.	No	300	2
										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	
											No	Yes
									HSP P.S.	No	130	0
										Yes	198	190

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

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.

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. 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 and 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 (Stegers-Jager et al. 2011; Eijsvogels et al. 2015), 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 with previous findings can be explained by previous investigations on academic dismissal making 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. Although progressing medical students' GPA is better under the higher-stakes assessment policy, the mimicked performance indicator 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. Secondly, 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 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 (Wolf and Smith 1995; Cole and Osterlind 2008). 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, Sanders, and Oreopoulos 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 similarly 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 (business administration), one course program in which both stakes and performance standard were adapted (medicine), and one course program in which the stakes, performance standard and resit standard were adjusted (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 will tell.

Thirdly, possible negative effects of raising the standards need to be considered as well. Negative consequences may include a lowered motivation for lifelong learning (Harlen and 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. In addition, since assessments are often unable to cover the full range of learning activities (Biggs 1996; UNESCO International Bureau of Education 2015), an increased focus on assessments may lower the time and energy devoted to the unassessed learning activities. Students' wellbeing also needs to be monitored. Higher standards may raise student stress-levels, which are associated with health problems (Glaser and Kiecolt-Glaser 2005) and lower academic performance (Akgun and Ciarrochi 2003). Finally, 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 and 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 and Hoff 2006). A question then could be which types of students exist in

terms of sensitivity to the characteristics of the assessment policy. Although students seem to view grades as either good or bad (Boatright-Horowitz and Arruda 2013), 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

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.

Acknowledgements

We would like to thank Peter Hermus for his patient help with the data screening and preparation, and Joran Jongerling for his helpful advice on the analyses.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Rob Kickert is a PhD student in the Department of Psychology, Education & Child Studies at Erasmus University Rotterdam, The Netherlands. His research interests include motivation, self-regulation, academic performance, and the possible consequences of different assessment policies in higher education.

Marieke Meeuwisse, PhD, is assistant professor of Education at the Erasmus University Rotterdam. Her main research interest is (ethnic) diversity in higher education, from the perspective of the learning environment, interaction, sense of belonging, motivation and academic success.

Karen M. Stegers-Jager, PhD, is assistant professor at the Institute of Medical Education Research Rotterdam, Erasmus MC, University Medical Centre Rotterdam. Her research interests include (ethnic and social) diversity, assessment, and selection and admission of medical students and residents.

Peter Prinzie, PhD, is Professor of Pedagogical Sciences at the Department of Psychology, Education & Child Studies, Erasmus University Rotterdam, The Netherlands. His research spans the field of developmental psychopathology, personality psychology, and developmental psychology.

Lidia R. Arends, PhD, is Professor of Methodology and Statistics at the Department of Psychology, Education & Child Studies, Erasmus University Rotterdam, The Netherlands. Besides, she is biostatistician at the Department of Biostatistics, Erasmus University Medical Center, Rotterdam, The Netherlands. Her areas of interest include research methods, (logistic) regression analysis, multilevel analysis, systematic reviews, and meta-analysis.

ORCID

Rob Kickert http://orcid.org/0000-0001-8584-869X
Marieke Meeuwisse http://orcid.org/0000-0002-4930-9581

Lidia R. Arends http://orcid.org/0000-0001-7111-752X
Peter Prinzie http://orcid.org/0000-0003-3441-7157
Karen M. Stegers-Jager http://orcid.org/0000-0003-2947-6099

References

Akgun, S., and J. Ciarrochi. 2003. "Learned Resourcefulness Moderates the Relationship between Academic Stress and Academic Performance." *Educational Psychology* 23 (3): 287–294. doi:10.1080/0144341032000060129.

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. doi:10.1080/03075079.2013.858684.

Attewell, P., S. Heil, and L. Reisel. 2011. "Competing Explanations of Undergraduate Noncompletion." *American Educational Research Journal* 48 (3): 536–559. doi:10.3102/0002831210392018.

Bandura, A. 1982. "Self-Efficacy Mechanism in Human Agency." *American Psychologist* 37 (2): 122–147. doi:10.1037/0003-066X.37.2.122.

Biggs, J. 1996. "Enhancing Teaching through Constructive Alignment." *Higher Education* 32 (3): 347–364. doi:10.1007/BF00138871.

Boatright-Horowitz, S. L., and C. Arruda. 2013. "College Students' Categorical Perceptions of Grades: It's Simply 'Good' vs. 'Bad.'" *Assessment & Evaluation in Higher Education* 38 (3): 253–259. doi:10.1080/02602938.2011.618877.

Chester, M. D. 2005. "Multiple Measures and High-Stakes Decisions: A Framework for Combining Measures." *Educational Measurement: Issues and Practice* 22 (2): 32–41. doi:10.1111/j.1745-3992.2003.tb00126.x.

Cohen, J. 1992. "A Power Primer." *Psychological Bulletin* 112 (1): 155–159. doi:10.1037/0033-2909.112.1.155.

Cole, J. S., and S. J. Osterlind. 2008. "Investigating Differences between Low- and High-Stakes Test Performance on a General Education Exam." *The Journal of General Education* 57 (2): 119–130.

De Koning, B. B., S. M. M. Loyens, R. M. J. P. Rikers, G. Smeets, and H. T. Van der Molen. 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. doi:10.1080/03075079.2012.754857.

Douglas, K. M., and R. J. Mislevy. 2010. "Estimating Classification Accuracy for Complex Decision Rules Based on Multiple Scores." *Journal of Educational and Behavioral Statistics* 35 (3): 280–306. doi:10.3102/1076998609346969.

Eijsvogels, T. M. H., R. Goorden, W. Van den Bosch, and M. T. E. Hopman. 2015. "The Binding Study Advice in Medical Education: A 2-Year Experience." *Perspectives on Medical Education* 4 (1): 39–42. doi:10.1007/s40037-015-0164-1.

Elikai, F., and P. W. Schuhmann. 2010. "An Examination of the Impact of Grading Policies on Students' Achievement." *Issues in Accounting Education* 25 (4): 677–693. doi:10.2308/iace.2010.25.4.677.

Field, A. 2013. *Discovering Statistics Using IBM SPSS Statistics*. London: SAGE.

Glaser, R., and J. K. Kiecolt-Glaser. 2005. "Stress-Induced Immune Dysfunction: Implications for Health." *Nature Reviews. Immunology* 5 (3): 243–251. doi:10.1038/nri1571.

Grabe, M. 1994. "Motivational Deficiencies When Multiple Examinations Are Allowed." *Contemporary Educational Psychology* 19 (1): 45–52. doi:10.1006/ceps.1994.1005.

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

Johnson, B. G., and H. P. Beck. 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. doi:10.1207/s15328023top1503_4.

Kickert, R., M. Meeuwisse, K. M. Stegers-Jager, G. V. Koppenol-Gonzalez, L. R. Arends, and P. Prinzie. 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. doi:10.1080/02602938.2019.1580674.

Kickert, R., K. M. Stegers-Jager, M. Meeuwisse, P. Prinzie, and L. R. Arends. 2018. "The Role of the Assessment Policy in the Relation between Learning and Performance." *Medical Education* 52 (3): 324–335. doi:10.1111/medu.13487.

Laursen, B., and E. Hoff. 2006. "Person-Centered and Variable-Centered Approaches to Longitudinal Data." *Merrill-Palmer Quarterly* 52 (3): 377–389.

Lindo, J. M., N. J. Sanders, and P. Oreopoulos. 2010. "Ability, Gender, and Performance Standards: Evidence from Academic Probation." *American Economic Journal: Applied Economics* 2 (2): 95–117. doi:10.1257/app.2.2.95.

Locke, E. A., and G. P. Latham. 2002. "Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey." *The American Psychologist* 57 (9): 705–717. doi:10.1037/0003-066X.57.9.705.

McManus, I. C. 1992. "Does Performance Improve When Candidates Resit a Postgraduate Examination?" *Medical Education* 26 (2): 157–162. doi:10.1111/j.1365-2923.1992.tb00142.x.

Olivier, J., and M. L. Bell. 2013. "Effect Sizes for 2 × 2 Contingency Tables." *PLoS ONE* 8 (3): e58777. doi:10.1371/journal.pone.0058777.

Pell, G., K. Boursicot, and T. Roberts. 2009. "The Trouble with Resits …." *Assessment & Evaluation in Higher Education* 34 (2): 243–251. doi:10.1080/02602930801955994.

Schippers, M., A. W. A. Scheepers, and J. B. Peterson. 2015. "A Scalable Goal-Setting Intervention Closes Both the Gender and Ethnic Minority Achievement Gap." SSRN Scholarly Paper ID 2647142. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2647142.

Sneyers, E., and K. De Witte. 2017. "The Effect of an Academic Dismissal Policy on Dropout, Graduation Rates and Student Satisfaction. Evidence from The Netherlands." *Studies in Higher Education* 42 (2): 354–389. doi:10.1080/03075079.2015.1049143.

Stegers-Jager, K. M., J. Cohen-Schotanus, T. A. W. Splinter, and A. P. N. Themmen. 2011. "Academic Dismissal Policy for Medical Students: Effect on Study Progress and Help-Seeking Behaviour." *Medical Education* 45 (10): 987–994. doi:10.1111/j.1365-2923.2011.04004.x.

Stegers-Jager, K. M., and A. P. N. Themmen. 2015. "Binding Study Advice: Effect of Raising the Standards?" *Perspectives on Medical Education* 4 (3): 160–162. doi:10.1007/s40037-015-0180-1.

UNESCO International Bureau of Education. 2015. "Student Learning Assessment and the Curriculum: Issues and Implications for Policy, Design and Implementation." *Current and Critical Issues in the Curriculum and Learning Series*. http://www.ibe.unesco.org/en/document/student-learning-assessment-and-curriculum-issues-and-implications-policy-design-and

Vossensteyn, H., A. Kottmann, B. Jongbloed, F. Kaiser, L. Cremonini, B. Stensaker, E. Hovdhaugen, and S. Wollscheid. 2015. "Dropout and Completion in Higher Education in Europe: Main Report." http://doc.utwente.nl/98513/

Wolf, L. F., and J. K. Smith. 1995. "The Consequence of Consequence: Motivation, Anxiety, and Test Performance." *Applied Measurement in Education* 8 (3): 227–242. doi:10.1207/s15324818ame0803_3.

Yocarini, I. E., S. Bouwmeester, G. Smeets, and L. R. Arends. 2018. "Systematic Comparison of Decision Accuracy of Complex Compensatory Decision Rules Combining Multiple Tests in a Higher Education Context." *Educational Measurement: Issues and Practice*, 37, 24–39. doi:10.1111/emip.12186.