

High stakes in the assessment arms race: a mixed-methods analysis of reducing assessment in an undergraduate psychology programme

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Pye, R. E. ORCID: https://orcid.org/0000-0002-1395-7763, Gnagi, L. K., Chuah, A. B. Q. and Chung, K. L. ORCID: https://orcid.org/0000-0003-0012-8752 (2025) High stakes in the assessment arms race: a mixed-methods analysis of reducing assessment in an undergraduate psychology programme. Psychology, Learning & Teaching. ISSN 1475-7257 doi: 10.1177/14757257251336295 Available at https://centaur.reading.ac.uk/122570/

It is advisable to refer to the publisher's version if you intend to cite from the work. See <u>Guidance on citing</u>.

To link to this article DOI: http://dx.doi.org/10.1177/14757257251336295

Publisher: Sage

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in



the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Reports

LEARNING & TEACHING

High Stakes in the Assessment Arms Race: A Mixed-Methods Analysis of Reducing Assessment in an Undergraduate Psychology Programme Psychology Learning & Teaching I-19 © The Author(s) 2025 ©

Article reuse guidelines: sagepub.com/journals-permissions DOI: 10.1177/14757257251336295 journals.sagepub.com/home/plj



Rachel Emma Pye

University of Reading, Reading, UK

Lynnette K Gnagi University of Reading Malaysia, Johor, Malaysia

Ainslie Bi Qi Chuah University of Reading Malaysia, Johor, Malaysia

Kai Li Chung D University of Nottingham Malaysia, Selangor Darul Ehsan, Malaysia

Abstract

Reducing assessment load has been identified as one practical way to allow space within programmes and modules for students to develop deeper learning practices and to increase student satisfaction. Yet despite the obvious benefits, both staff and students report preferring to maintain a high assessment load as a way to ensure student attention (staff), or to manage risk (students). This mixed-methods study looked at the reduction of the number of assessments across final-year optional modules in the international branch campus of a UK university psychology programme on module grades and student perceptions. We found that module grades did not increase following reduction. Students reported anxiety about single-assessment module regimes, regardless of their experience of assessment reduction. Students overwhelmingly preferred two assessments per module, interestingly on the grounds of fairness from a diverse assessment portfolio. We suggest that a simple reduction in the number of assessments isn't itself sufficient to meet the broad aims of slow scholarship, but that programme teams could consider how better as well as fewer assessments and the perceptions of workload might be more important to tackle.

Corresponding Author:

Rachel Emma Pye, School of Psychology & Clinical Language Sciences, University of Reading, Reading, RG6 6AL, UK. Email: rachel.pye@reading.ac.uk

Keywords

Assessment load, anxiety and risk, mixed-methods, fairness in assessment, workload

Introduction

Assessment is a key element of all higher education programmes, with research focussing on good design, constructive alignment, and how format affects student learning (see Pereira et al., 2016, for review). Despite increased attention to assessment design (e.g. Bearman et al., 2017), there is little work considering the practical impacts of programme-level assessment reduction, which reflects Evans et al. (2021)'s assertion that pedagogical research often fails to consider the local practitioner context. Across theoretical perspectives, contextual factors such as course structure are proposed to interact with individual-level factors such as perceptions, engagement, or approaches to learning. Biggs (1989) introduced the 3P model, further developed by Prosser and Trigwell (1999) and more recently considered part of the Student Approaches to Learning (SAL) framework (see Zusho, 2017) explaining how Presage factors (such as course structure, students' preceding knowledge, etc.) interact with Process factors, such as student perceptions and approaches, and Product factors, which reflect student learning outcomes and are often measured by summative assessment performance. In this report, we used a mixed-methods approach to explore a small-scale trial of assessment reduction (Presage factor) in a UK psychology programme, using its smaller sister campus in Malaysia to qualitatively explore student attitudes to assessment (Process factors) alongside a quantitative exploration of module marks (Product factors) during the period of assessment change. Such an approach allows us to explore how psychology programme design teams could reduce student (and staff) workload, whilst exploring students' concurrent perceptions of assessment, workload, and learning.

Recent attention has been paid to the impact of assessment load on learning, which Jessop and Tomas (2017) defined as the number and diversity of assessments experienced by students throughout their programme of study. It has been noted that despite calls to control assessment number and type (e.g. Harland et al., 2015; Jessop & Tomas, 2017), academics still tend to see assessment as a means to ensure engagement with their module / area (O'Neill, 2019). Harland et al. (2015) termed this an 'assessment arms race', in their call for academic staff to move towards a 'slow scholarship' (Hartman and Darab, 2012) approach to teaching and learning. Harland et al. (2015) interviewed staff and students at a New Zealand university to investigate the impact of the introduction of modularisation. They found that students reported wanting multiple low-stakes summative assessments to manage risk, despite their recognition that these assessment regimes led to surface learning and sub-optimal quality of work. Harland et al. (2015) called for staff to consider reducing assessment load as a way to allow space for deeper and slower learning, in a context where students were completing assessments almost daily; in other words, to reduce assessment to encourage a culture of slow scholarship. Harland and Wald (2021) more recently discussed that slow scholarship through reduction of assessment takes a long time, but one possible method to reduce assessment would be to remove smaller weighted assessment piece-by-piece whilst monitoring change on learning, the approach explored in this report.

Lizzio et al. (2002) proposed that the best place for programme designers to start when attempting to raise student attainment and satisfaction is by addressing workload and assessment, given that these factors consistently show a robust relationship to approaches to learning and are less resource-intensive than changes made to teaching, which require investment in staff training, development, etc. However, as Galvez-Bravo (2016) noted, little research has looked at the quantitative impact of overassessment on student achievement and satisfaction, rather than perceptions of the same. They found a trend, though non-significant, for higher student evaluation of modules with a higher number of summative assessments compared to those with lower numbers of summative assessments, and a significant positive correlation between student evaluation and overall module marks across all modules, perhaps implying that small-scale low-stakes assessment leads to higher grades and related satisfaction. Interpreting assessment load impacts on modules is a challenge in naturally occurring data sets such as those used by Galvez-Bravo (2016). It is notable that there is still a paucity of research on the effects of assessment strategy on achievement using quantitative approaches. Therefore, in this report, we aimed to utilise both quantitative and qualitative

approaches. Therefore, in this report, we aimed to utilise both quantitative and qualitative approaches in a mixed-methods design, which enables a more nuanced exploration of the relationships between assessment and students' attainment.

The Current Report

As part of an institution-led review, the psychology programmes team identified that our students experienced a 'high' number of assessments according to Jessop and Tomas's (2017) Transforming the Experience of Students Through Assessment categorisation, with a high weighting of marks from exams (see Appendix A). In the final undergraduate year a wide variety of small (10-credit) optional modules were available, which all followed the same assessment pattern of 25% coursework and 75% unseen exam. As the psychology programme is also delivered on University of Reading's international branch campus in Malaysia (UoRM), which enrols ~ 20 students (compared to ~ 250 at UoR) annually, it was agreed that a reduction in assessment in final-year optional modules would be trialled at UoRM in academic year 2020/21. Successful implementation would lead to the same on the larger campus. Staff at UoRM therefore decreased assessment from two unequally balanced assessments to one 100% assessment in all optional modules, with type of assessment determined by the module convenor as appropriate to the module content. Two compulsory modules remained unchanged across both campuses, enabling marks on these modules to be used to check comparability of cohorts across the period of reduction. This provided a unique opportunity to explore the reduction of assessment on student summative grades (Product) and student perceptions (Process) on a small cohort. Additionally, the fact that these constituted 100% of the optional module assessments also allowed investigation of students' concerns around 'high-risk' assessment. This therefore provided a more controlled opportunity to explore assessment reduction and student marks quantitatively, similar to the analysis by Galvez-Bravo (2016), and used mixed methodology to enable appropriate comparisons with qualitatively approached analyses of similar topics.

Given that slow scholarship suggests that fewer assessments allow more space for deeper learning approaches, we hypothesised that the reduction of assessment in these high-stakes modules, from 14 to eight pieces across the final year, would lead to better performance in module assessment. We looked at the module grades of all UoRM Year 3 students for the 2 years preceding and 2 years following the change in assessment structure in academic year 2020/21, from 2018/19 to 2021/2, including grades in compulsory modules with unchanged assessment structures as a comparison. Students on both the UK and the Malaysia campus also participated in a small series of focus groups, with final-year students included to compare perceptions of the old and new assessment structures, and second-year students on both campuses to explore perceptions of their upcoming assessment structures. Through qualitative analyses, we expected students who had completed one assessment per module (the 'Year 3' students at UoRM) to be more positive about their perception of workload, having experienced fewer assessments than their contemporaries in Year 3 at UoR. We expected Year 2 students on both campuses to be concerned with assessment in Year 3, but to recognise the possible advantages of fewer assessments during this period. We were interested to see whether approaches to learning were different when assessment load was reduced, as suggested by slow scholarship aims. Therefore, using a mixed-methods approach we explored:

- 1. Whether summative module grades would improve following the reduction in number of assessments in optional modules on the smaller Malaysia campus, using quantitative analyses, and;
- 2. Students' experiences and perceptions of their programme assessment structures, and whether these were more positive when students had fewer assessments, using qualitative analyses.

Method

Participants and Design

We used a mixed-methods research design (Creswell & Plano Clark, 2018). For this quasiexperimental between-subjects design, quantitative data from four cohorts of Year 3 students in the academic years 2018/19 to 2021/22 at UoRM were analysed to explore differences in summative grades for two cohorts preceding and two cohorts following the move to 100% single assessment in final-year optional modules in 2020/21. Module marks from the modules with reduced assessment were analysed, as were overall marks from Year 2 and two compulsory modules from Year 3, which were unchanged in content and assessment structure across this period, to enable comparison across cohorts with regard to performance. Qualitative data were collected through focus groups with Year 2 and Year 3 students on both campuses at the end of the academic year 2020/1, with those in Year 3 at UoRM the first cohort to experience the reduced-assessment structure, as detailed below, to explore potential differences in student perceptions around assessment. See Figure 1 below.

Focus Groups / Interviews

A series of focus groups run and facilitated by five Year 2 student partners on both campuses looked at student attitudes and feelings towards assessment structure. One focus group with Year 2 students



Figure 1. Schematic illustration of the assessment structure on UoR and UoRM campuses pre- and post-optional module assessment reduction in 2020/21. Note that focus groups were run at the end of the 2020/21 academic year.

(n=3 UoR, n=5 UoRM), and one focus group of Year 3 students (n=4 UoR, n=5 UoRM) was run on each campus, over Microsoft Teams, facilitated by students registered on that campus.

An interview schedule (see Appendix B) was created collaboratively between the authors, two of whom were senior student partners. The main topics explored were around student perceptions and feelings of assessment structure, specifically around 100% single assessment of small modules, and how such assessment structures impacted their approach to work within the module, and decision-making around optional module choice.

Two student partners facilitated each focus group session, with one being the note-taker. The Microsoft Teams sessions were recorded and later transcribed. The discussions lasted from 30 to 75 min. At the closing of the session, student partners sent participants a link to a textbox where they could share any afterthoughts.

Ethics

This study was given a favourable opinion for conduct from the University of Reading School of Psychology and Clinical Language Sciences Research Ethics Committee (reference 2021-091-RP) and was conducted in accordance with the principles stated in the Declaration of Helsinki. Informed consent was obtained for focus group participants. Recordings of focus groups were destroyed once transcribed, in keeping with the University of Reading's policy on data management.

Data Analysis

Quantitative Analysis. Mean module marks (out of 100) at UoRM were analysed for unchanged compulsory and changed optional modules in year 3, as well as year 2 and year 3 weighted average marks, across four cohorts, two which directly preceded the 2020/21 change in assessment structure (2018/9, n = 7; 2019/20, n = 16), and two which directly followed the assessment structure change (2020/21, n = 19 and 2021/22, n = 22).

Qualitative Analysis. We used Braun and Clarke's (2012) reflexive thematic analysis approach which follows a six-phase process. We first familiarised ourselves with the data, reading the transcripts actively whilst noting down items of interest. We then generated a list of codes using labels that captured the richness as well as nuances of the data; the coding was done manually a couple of times inclusively, comprehensively, and systematically to ensure each data item was given equal attention. Themes were generated by organising and clustering similar codes. We also considered the relationships between themes to develop a central idea that relates to our research question. Next, we reviewed the potential themes, whilst making sure that there were enough meaningful data to support the themes generated and none were overly fragmented. We then defined and named generated themes. Finally, we produced the write-up below, identifying examples of data to illustrate our themes in an order that tells the overall story of our analysis.

Results

Quantitative Analysis: Module Data

An ANOVA was used to explore module grade differences in two cohorts pre- and two cohorts post-introduction of the 100% single-assessment structure in the academic year 2020/21. There were no significant differences between the four cohorts on year 2 overall average mark

Table 1. Cohort Analysis of Year 2 for the two Cohorts Preceding Ass	and Year 3 Ov essment Redu	erall Year Meaı ction and two	ns and Stand Cohorts Fo	ard Deviatio Ilowing Asse	ns for UoRN ssment Rec	1 Students, a Juction.	ind Compul	sory Year 3	Modules (ou	t of 100)
			Year 2 av	erage	Year 3 av	erage	40c com	pulsory	20c comp	ulsory
2 assessments in option modules	2018/19	(n = 7)	64.88	(10.53)	59.29	(7.54)	60.0	(8.58)	54.57	(8.34)
	2019/20	(n = 16)	62.19	(7.13)	60.81	(2.01)	61.06	(7.59)	57.5	(8.77)
l assessment in option modules	2020/21	(n = 19)	61.35	(7.39)	58.47	(5.38)	60.95	(8.98)	58.0	(6.77)
	2021/22	(n = 22)	64.29	(7.16)	62.5	(7.67)	64.23	(8.72)	65.91 ^a	(7.12)

^aSignificantly different module grade to earlier cohorts. All other measures show no significant cohort differences.

(F(3,60) = .707, p = .551), indicating cohorts were performing similarly across the period. Of the two compulsory modules taken by all students, there was no significant difference across years for the 40-credit module (F(3,60) = .791, p = .501) but there was a significant difference across cohorts for the 20-credit module (F(3,60) = 6.603, p = .001). Post hoc analysis indicated the differences were driven by a higher mark in this module in the 2021/2 cohort compared to all others (see Table 1 below).

All optional modules had reduced assessment from 2020/21. Three optional modules ran for all four cohorts (2018, 2019, 2020, 2021), and two optional modules ran over 3 of the 4 years (see Table 2 below). Analysis of variance demonstrated there were no significant differences between cohorts for Option 1 (F(2,54) = .772, p = .467), Option 2 (F(3,44) = .252, p = .86), or Option 5 (F(2,26) = .363, p = .699). Significant differences were found for Option 3 (F(3,57) = 3.385, p = .024) and Option 4 (F(3,54) = 3.741, p = .016). Post hoc analyses showed significantly lower module grades in 2020 compared to 2019 cohorts (i.e. immediately pre-100% assessment and post-100% assessment) for both Option 3 (p = .049) and Option 4 (p = .026). Unplanned exploratory analyses indicated that in the two-assessment model both these modules had lower coursework marks, and higher exam marks, meaning that removal of the exam component led to lower module marks overall (see Figure 2 below).

Given the small sample size, it is unclear whether the null results across the four cohorts on three of the five reduced-assessment optional modules can be considered to demonstrate no difference, or insufficient evidence of difference. Using Lakens et al.'s (2018) approach, we therefore used 90% confidence intervals to identify if module grades pre- and post-assessment reduction demonstrated equivalence. We determined that the Smallest Effect Size of Interest (SESOI) would equate to a difference of 5 marks, which is equivalent to half a grade boundary for final classification in the UK system. We calculated the Cohen's d based on the actual Year 2 overall mark and the overall mark 5 marks higher, using the online effect size calculator at psychometrica.de, which gave a d of 0.613. We then plotted the 90%CI for the overall Year 3 marks of students pre



Figure 2. Comparison of summative assessment mean marks (out of 100) in 2019/20 (2 assessments per module) and 2020/21 (one assessment per module) for essay-based optional modules. Note: Error bars represent standard errors.

Table 2. Cohort	Analysis of	Optional	Part 3 M	odule Mea	n Marks (out of 10	o).									
		Option	_		Option	2		Option :	3		Option	4		Option	5	
Assessment type		Case re	port		Researc	h report		Essay			Reflectiv	/e report		Poster		
25% coursework, 75% exam 100% coursework ^a Significantly differ.	2018/19 2019/20 2020/21 2021/22 ant module	61.56 57.95 58.77 88.77 87ade to	(6.45) (8.58) (11.01) earlier ar	(n = 16) (n = 19) (n = 22) nd later col	58.2 60.67 58.69 61.06 horts. All	(12.11) (8.02) (7.76) (10.46) other mo	(n = 5) (n = 9) (n = 16) (n = 18) odule meas	57.43 63.31 56.95 ^a 64.27 sures shov	(6.27) (4.7) (9.1) (1023) w no signif	(n = 7) (n = 13) (n = 19) (n = 22) ficant coho	54.5 63.19 54.56 ^a 60.9 rrt differer	(18) (8.37) (7.5) (6.99) nces.	(n = 4) (n = 16) (n = 18) (n = 20)	58.14 55.67 60.13	(10.85) (10.07) (10.87)	(n = 7) (n = 6) (n = 16)

Psychology Learning & Teaching 0(0)



Figure 3. 90% confidence interval test of equivalence based in a SESOI (smallest effect size of interest) of 5 marks (d = .613) for year 3 overall marks (OWA) and modules with reduced assessment before and following reduction.

(i.e. 2018 + 2019, n = 23) and post (i.e. 2020 + 2021, n = 41) reduction. Figure 3 (below) demonstrates that Year 3 overall marks, and some but not all optional modules, are within the SESOI boundary, thus partially demonstrating equivalence.

Qualitative Analysis: Themes

We generated two themes: Assessment Weight and Assessment Variety. Four sub-themes emerged: Fairness, Risk and Anxiety, Workload, and Approach to Learning. The links between these themes and sub-themes can be seen in Figure 4 below. Sub-themes are discussed in detail below.

Fairness. Fairness was an important element of students' perceptions and feelings towards both assessment weight and assessment variety. Having multiple diverse assessments was perceived to be a fairer way of assessing learning for those with different strengths and preferences.

So, having like a variety of assessments uh for a single module would be good in that sense, where it's it gives everybody equal chances to shine in the places that they that they are best at (Participant 10, Year 2, UoRM)

I would love for it to be 100% coursework. But yeah, some people do prefer exam, so I'd say like at least 50:50 (Participant E, Year 2, UoR)

The preference for some types of assessments also relates to a sense of control, in that having time and resources enables students to have greater control over how they complete an assessment, particularly if it is a written essay.



Figure 4. Diagram of themes and sub-themes.

I think when it's essay based, in my head, it just makes it fairer for it to be coursework because you're doing the same thing, but you're just given the time and resources to do it to a good quality. (Participant E, Year 2, UoR)

Risk and Anxiety. When asked about how they felt about having only one assessment per module, participants expressed anxiety and showed a preference for having multiple assessments. Overall, participants felt that if they '*mess up a piece of coursework*', '*there*'s no chance to redeem yourself'. Having multiple assessment is one way to mitigate this risk.

I find it very stressful if module is 100% coursework and is one piece of work. (Participant G, Year 2, UoR)

But to me it was also quite stressful when I was doing assignment because again it was 100% weightage, so it was a do or die moment. So I would have appreciated, I would have in a sense rather have... a balance in weightage maybe like a 30% or 40% in our assignment and you know a good chunk of it, put it in our exams. (Participant 1, Year 3, UoRM)

Say your 25% coursework wasn't up to standard, then you know you've got the 75% exam to sort of fall back on. (Participant H, Year 3, UoR)

However, their feelings were also influenced by the type of assessment. One participant commented that they would be more confident in taking a 100% take-home examination as opposed to a 75% unseen examination. This preference is driven by circumstances in which participants feel they have no control over, such as examination conditions:

So I think I would like much more confidence in, in having like a take home assessment that is worth 100% as compared to like a having to have like a written exam which is worth like 75% because it's like what if on the day itself during that written exam like I have like, you know, there are a lot of conditions that I cannot prepare myself for, or get like, prepare myself for like like nervousness or you know, (Participant 10, Year 2, UoRM)

Workload. Paradoxically, students were anxious about the high-stakes single-assessment model but were equally concerned about workload management difficulties associated with having more than one assessment. This led students to feel overwhelmed, which in turn led to increased anxiety that had an impact on how they perceived the assessment models.

I ended up putting too much effort into something that was like closer, like courseworks. And then I was neglecting um the dissertation, which is obviously like really important... What do I spend more time on and like it's kind of hard to divide your time. (Participant D, Year 3, UoR)

That leads to quite a big sort of pile up of deadlines and exams, and everything becomes a bit more overwhelming because of that. (Participant G, Year 2, UoR)

Some students used the assessment weighting to guide their decisions about importance, but tended to look at this at a module rather than programme level, for example believing a 75% exam of a 10 credit module was worth more than 50% of a 20 credit module.

I found because uh the coursework was only 25 percent, I was... I didn't take it seriously as I did previous years 'cause I knew it wasn't that much, it's fine. (Participant B, Year 3, UoR)

I think when um in second year it was what, like 20 uh 50 percent coursework and 50 percent exam. So that jump to like 75 percent exam was kind of terrifying. (Participant A, Year 3, UoR)

Workload in terms of deadline distribution was stressful, and bunched deadlines were perceived as unfair.

That's not fair though because then you end up being mediocre in like across all of them as opposed to doing really well in and. I just thought that was really unfair because we had like three. (Participant C, Year 3, UoR)

Approach to Learning. One assessment was seen as inadequate to capture all the module learning outcomes; interestingly for some participants exams were seen as a better way to ensure content was covered in breadth and depth compared to coursework.

But yeah, uh, and on top of that, I think that the scope of our content that that we are tested on is very narrowed lah when you have only one assessment. (Participant 2, Year 3, UoRM)

And for me I think I prefer examination overall, uh, because... it actually forces me to study, to understand the concept and that's the thing. (Participant 8, Year 2, UoRM) whereas with exams you can kind of choose what you want, which I think is just a lot easier 'cause you have a bit more control. (Participant C, Year 3, UoR)

Prior to the Curriculum Framework Review, continuous assessment questions (CAQs) were a major component of the assessment for year 1 and 2 students. These were typically in the format of a series of multiple-choice questions (MCQs) administered on a weekly basis, designed to provide ongoing analysis of students' learning performance. Interestingly, participants seemed to perceive CAQs as a measure of content recall, but the continuous nature of the assessment gives them an opportunity to improve their performance:

I think like, the reason why I think CAQ would be a very good option because... It's either you, you know it or you don't kind of thing. (Participant 10, Year 2, UoRM)

Uhh, compared to writing an essay, because I think there are some questions that I also prefer, like those CAQ formats.... so that I get a chance to improve? (Participant 9, Year 2, UoRM)

When directly asked about formative assessment, some students did not know what this meant, and some students believed that they had not been given appropriate opportunity to do any after the first year.

I think I had only a few right in year one other than that no right? (Participant 2, Year 3, UoRM)

Yeah, I think formative might actually like... be more helpful for especially first year. I mean obviously second and third year you're used to at that point, (Participant A, Year 3, UoR)

I wish we had um a formative example like of how to do a poster assignments as well. (Participant B, Year 3, UoR)

Others however recognised the use and utility of in-class formative assessment:

It does help when it comes to like, um, writing stuff for essays or understanding the topic. (Participant 3, Year 3, UoRM)

Participants had anticipated certain types of assessment, such as essays and CAQs, occurring in the final year based on assessment types in earlier years. However, not all options used these methods of assessment, with a wider variety available.

None of my modules ask for an essay this year and I was like that's all I've been taught. (Participant B, Year 3, UoR)

I also missed the CAQs 'cause I think in first and second year they kind of like you get so used to the CAQs and then it's kind of a shock when you don't have CAQs for most modules in third year. (Participant A, Year 3, UoR)

Again, some students but not all considered CAQs as a method for spreading risk and increasing marks:

...why not have the CAQ format again?... because OK, at the end of the day, what I want for my year 3 assessment is for me to score, to have the highest chance of scoring the most marks as possible, right? That's, I mean, as practical and as materialistic as that sounds, I want to get my highest mark possible. So, with that being said, I would think that to spread out my chances would be a better, safer choice for me. (Participant 10, Year 2, UoRM)

... when I pull out my whole transcript, I noticed my CAQs brought down all my marks. Whereas, my essays boost it up. So CAQs is like, to me MCQs, it's very dangerous in general. (Participant 6, Year 2, UoRM)

Whilst assessment is a main source of anxiety throughout a students' course of study, the assessment type does not seem to affect module choice, with many preferring to pick an option in which they find the content interesting. This was particularly relevant to the UoX cohort, where students get a wider range of optional module choices.

So in that case I chose one business module which is marketing and that's the one that interests me, but that one has exam and so I don't really mind whether it's a full assignment or assignment plus exam. (Participant 8, Year 2, UoXRM)

But I would go with content more just because, coursework and exam, like I'm more likely to put more work in if I'm enjoying the content. (Participant F, Year 2, UoR)

Discussion

Using a mixed-methods approach we explored whether reduction in assessment, an environmental Presage factor, was associated with better student performance as measured by overall year 3 marks and individual module grades, a Product factor. Following the ideas of slow scholarship (see Harland et al., 2015), we hypothesised that reduced assessment might lead to improved marks because the additional space created enables deeper learning. We were also interested in exploring whether previous research indicating that students have a preference for multiple assessment held true once assessment had been reduced.

The quantitative data analysis indicated that there were no significant improvements in overall year 3 marks, confirmed using 90% CI equivalence indicators, following the reduction in assessment (2020 and 2021 cohorts combined) compared to those preceding any change (2018 and 2019 cohorts combined). Indeed, there was evidence for a decrease in some module marks immediately following the change for two modules, and some modules had data which was insufficient to draw conclusions regarding difference or equivalence. Unplanned exploratory analysis indicated that in the two optional modules with a lower mark following assessment reduction, the exam marks had been higher than coursework marks under the two-assessment model, and the coursework remained unchanged as sole assessment in 2020. From this pattern of data, we tentatively conclude that simple reduction of assessment does not lead to better grades as a result of creating space for deeper learning and slower scholarship. At best, the equivalence tests demonstrate that some module marks, and the year 3 overall mark, were not negatively impacted by this reduction, which indicates that simple reduction in the number of assessment could reduce workload for students and staff without detriment to grades. The ANOVA and equivalence tests also indicate that some module marks were lower following the reduction needs to be considered carefully.

The qualitative analysis of focus groups clearly indicated that students prefer multiple assessments, supporting the previous findings of Harland et al. (2015), and Galvez-Bravo (2016). The reasons for this were slightly different to those discussed in Harland et al. (2015), whose students reported being aware of the unsatisfactory nature of multiple assessments, but felt that these managed risk. Our students also wanted more than one assessment, with two 50% assessments the preferred model, partly due to spreading risk, but also partly due to a sense of fairness for colleagues with regard to diversity of assessment. Additionally, students at UoXY still struggled to manage their workload in the final year, even though the number of assessments had been reduced. All of these findings relate to Tomas and Jessop's (2019) proposal that a variety of assessment is needed, but too much causes cognitive overload for students. We did not directly explore the paradox here, though SAL frameworks, supported by empirical evidence such as Lizzio et al. (2002) emphasise it is the *perception* of workload rather than workload itself which is related to student dissatisfaction. It is interesting that our participants largely stated they would prefer two assessments rather than one, but also stated they liked the weekly CAQ format, which is clearly an assessment in the image of Harland et al.'s 'pedagogy of control'.

In summary, it was noticeable that the students who had been through the reduced-assessment structure (Year 3 UoXY) felt very similarly to those who hadn't (Year 3 UoX), and those who were going to have reduced assessment (Year 2 on both campuses). We therefore found no evidence to support our assumption that a reduction in assessment would lead to more positive perceptions of workload, i.e. a change in Process. We also found no evidence of improved performance in assessment, i.e. no change in Product.

Methodological Considerations and Future Directions

We acknowledge that there is an assumption in this study that deeper learning will lead to higher grades for students, which might be due to an overly simplified interpretation of slow scholarship, which emphasise the importance of student autonomy and changed staff-student dynamics (see Harland and Wald, 2021) and sees assessment as an outcome of a 'pedagogy of control'. Whilst slow scholarship espouses a reduction in assessment load, especially of frequent and low-weighted varieties, there is a literature indicating that such assessment methods, requiring retrieval of learning during teaching periods, can lead to improved learning outcomes. A recent special issue of PLAT (see Kubik et al., 2021) explored moderating factors such as learner characteristics and feedback on the effect of retrieval practice on learning. Den Boer et al.'s (2021) study indicated that this type of frequent, spaced testing led to higher student motivation for the regular tests, but didn't impact the final grade in the exam, drawing the conclusion that these could be used as formative rather than summative activities. There is a tension between these approaches: assessment as control vs assessment for learning, and practitioners should consider these tensions when designing assessment for modules, and across programmes.

We also acknowledge the inherent danger of over-interpreting changes in naturally occurring data sets such as module marks over time, especially given that cohorts were small. We did attempt to control for some Process / individual factors by confirming that Year 2 and compulsory modules were similar across cohorts, which should capture some motivation / aptitude cohort-level differences. However, future research could more directly address potential alternative explanatory individual learner factors in a more controlled manner to better explore the effects of assessment reduction on student attainment and perceptions.

It is also notable that we are comparing module marks across cohorts who were variably affected by the COVID-19 pandemic, which led to long lockdown periods in Malaysia with students experiencing online teaching and learning from February 2020 to December 2021 which could have impacted student experience of assessment. As with many UK HEIs, a 'safety-net' approach to assessment was introduced in academic year 2019/20, whereby students were awarded the higher mark of either pre-lockdown work, or the total of pre- and mid-lockdown work. Analysis of our cohorts shows that there were no differences between cohorts with weighted average marks for year 2 or year 3, and the only differences seen across cohorts was for one compulsory final-year module with 2021/22 students performing better that the other cohorts. We are therefore relatively confident that COVID-19 did not affect the quantitative data interpretations presented here. However, it is possible that student anxiety and perceptions of workload were affected by the pandemic: from studies on the effect of the pandemic on student mental health (e.g. Chen and Lucock, 2022), it seems likely that these were increased and therefore could potentially mask any expected improved perceptions of workload following assessment reduction.

Future research could more explicitly explore the impact of broad environmental Presage factors on downstream Product factors through longitudinal large-scale studies which also account for individual-level Process factors. Whilst the interaction between classroom context and individual factors has been well-researched across the theoretical frameworks of self-regulated learning, SAL and engagement (see Zusho, 2017), much of this work has focussed on individual factors. Student perceptions of their learning environment seem to be a key factor in the adoption of deep learning approaches (Guo et al., 2022), although Biggs et al. (2001) maintained that student factors, teaching context and approaches to learning interact to form a dynamic system, implying that changes at programme level (Presage) should impact the Process and Product factors.

Recommendations for Educational Practise

The quantitative data indicate no difference in Year 3 marks and for most individual modules following assessment reduction. Indeed, for modules that merely 'dropped' an exam rather than redesigning assessment, marks were lower. The exams required students to know and remember material across the module, whereas essay questions in both modules focussed on a single-topic. This was addressed for the following cohort, and the marks showed a trend to return to a similar mean to that of earlier (pre-reduction) cohorts. We therefore tentatively interpret these findings as indicating that reducing assessment needs to be managed proactively; we would also encourage practitioners, given students' comments during the focus groups about the benefits of exams, to ensure synoptic coverage of material across assessments.

In focus groups students reported preferences for two assessments per module, with some differences in what kind of assessment should be used, to spread risk and increase fairness, with the caveat that they had experience of these assessment types. Given that we noted that students thought of assessment weight at a module rather than programme level, it could be possible for programme designers to reduce both assessment number and student anxiety by having a smaller number of larger credit-weighting modules (e.g. 6×20 credits vs 12×10 credits within the UK system of 1 academic year = 120 credits), each of 2 or 3 assessments.

Module assessment didn't factor heavily in module choice, although some students did report that assessment type somewhat impacted their approach to learning, with some saying that exams lead to better learning and others perceiving coursework to do so, reinforcing Tomas and Jessop's (2019) suggestion that diversity of assessment type is preferential, but that too much variety leads to a burden on students' cognitive load.

We conclude that simple reductions in assessment do not immediately lead to improvements in marks, or to improved perceptions of workload. Better assessment design plus fewer assessments could lead to higher student satisfaction through changed perceptions of workload if managed appropriately, but the effects of Presage factors such as assessment load on Product factors such as module marks are distal not proximal, and changing Process factors such as student perceptions

and approaches to learning are likely to be more impactful in the aim to enable deeper learning and better outcomes for our students.

Acknowledgements

The authors would like to acknowledge the contributions of our student partners on both UK and Malaysia campuses, who facilitated and transcribed the focus groups.

Disclosure Statement

The authors report there are no competing interests to declare.

Data Sharing Statement

The data that support the findings of this study are available from the corresponding author, RP, upon reasonable request.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by internal funding from the School of Psychology & Clinical Language Sciences Teaching & Learning Enhancement Projects Fund.

ORCID iD

Rachel Emma Pye D https://orcid.org/0000-0002-1395-7763 Kai Li Chung D https://orcid.org/0000-0003-0012-8752

References

- Bearman, M., Dawson, P., Bennett, S., Hall, M., Molloy, E., Boud, D., & Joughin, G. (2017). How university teachers design assessments: A cross-disciplinary study. *Higher Education*, 74(1), 49–64. https://doi.org/ 10.1007/s10734-016-0027-7
- Biggs, J. B. (1989). Approaches to the enhancement of tertiary teaching. *Higher Education Research and Development*, 8(1), 7–25. https://doi.org/10.1080/0729436890080102
- Biggs, J. B., Kember, D., & Leung, D. Y. (2001). The revised two-factor study process questionnaire: R-SPQ-2F. British Journal of Educational Psychology, 71(1), 133–149. https://doi.org/10.1348/ 000709901158433
- Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), APA Handbook of research methods in psychology, vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological (pp. 57–71). American Psychological Association.
- Chen, T., & Lucock, M. (2022). The mental health of university students during the COVID-19 pandemic: An online survey in the UK. *PLoS ONE*, 17(1), e0262562. https://doi.org/10.1371/journal.pone.0262562
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed). Sage.
- Den Boer, A. W. J. P., Verkoeijin, P. P. J. L., & Heijltjes, A. E. G. (2021). Comparing formative and summative cumulative assessment: Two field experiments in an applied university engineering course. *Psychology Learning and Teaching*, 20(1), 128–143. https://doi.org/10.1177/147572570971946
- Evans, C., Howson, C. K., Forsythe, A., & Edwards, C. (2021). What constitutes high quality higher education pedagogical research? Assessment & Evaluation in Higher Education, 46(4), 525–546. https://doi.org/10. 1080/02602938.2020.1790500

- Galvez-Bravo, L. (2016). Overassessment in higher education: Does less mean more? *Innovations in Practice*, 10(2), 93–102. https://doi.org/10.24377/LJMU.iip.vol10iss2article94
- Guo, J. P., Yang, L. Y., Zhang, J., & Gan, Y. J. (2022). Academic self-concept, perceptions of the learning environment, engagement, and learning outcomes of university students: Relationships and causal ordering. *Higher Education*, 83(4), 809–828. https://doi.org/10.1007/s10734-021-00705-8
- Harland, T., McLean, A., Wass, R., Miller, E., & Sim, K. N. (2015). An assessment arms race and its fallout: High-stakes grading and the case for slow scholarship. *Assessment & Evaluation in Higher Education*, 40(4), 528–541. https://doi.org/10.1080/02602938.2014.931927
- Harland, T. A., & Wald, N. (2021). The assessment arms race and the evolution of a university's assessment practices. Assessment & Evaluation in Higher Education, 46(1), 105–117. https://doi.org/10.1080/ 02602938.2020.1745753
- Hartman, Y., & Darab, S. (2012). A call for slow scholarship: A case study on the intensification of academic life and its implications for pedagogy. *The Review of Education, Pedagogy, and Cultural Studies*, 34(1–2), 49–60. https://doi.org/10.1080/10714413.2012.643740
- Jessop, T., & Tomas, C. (2017). The implications of programme assessment patterns for student learning. Assessment & Evaluation in Higher Education, 42(6), 990–999. https://doi.org/10.1080/02602938.2016.1217501
- Kubik, V., Gaschler, R., & Hausman, H. (2021). PLAT 20(1) 2021: Enhancing student learning in research and educational practice: The power of retrieval practice and feedback. *Psychology Learning and Teaching*, 20(1), 1–20. https://doi.org/10.1177/1475725720976462
- Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. Advances in Methods and Practices in Psychological Science, 1(2), 259–269. https://doi.org/10.1177/ 2515245918770963
- Lizzio, A., Wilson, K., & Simons, R. (2002). University students' perceptions of the learning environment and academic outcomes: Implications for theory and practice. *Studies in Higher Education*, 27(1), 27–52. https://doi.org/10.1080/03075070120099359
- O'Neill, G. M. (2019). Why don't we want to reduce assessment? *AISHE-J: The All Ireland Journal of Teaching and Learning in Higher Education*, 11(2), 1–7. https://doi.org/10.62707/aishej.v11i2.415
- Pereira, D., Assunção Flores, M., & Niklasson, L. (2016). Assessment revisited: A review of research in assessment and evaluation in higher education. Assessment & Evaluation in Higher Education, 41(7), 1008–1032. https://doi.org/10.1080/02602938.2015.1055233
- Prosser, M., & Trigwell, K. (1999). Understanding learning and teaching: The experience in higher education. McGraw-Hill Education.
- Tomas, C., & Jessop, T. (2019). Struggling and juggling: A comparative analysis of assessment loads across different universities. Assessment & Evaluation in Higher Education, 44(1), 1–10. https://doi.org/10.1080/ 02602938.2018.1463355
- Zusho, A. (2017). Toward an integrated model of student learning in the college classroom. *Educational Psychology Review*, 29(2), 301–324. https://doi.org/10.1007/s10648-017-9408-4

Author Biographies

Dr Rachel Pye is an associate professor in developmental psychology at the University of Reading, UK. Between 2017 and 2022, she was seconded to the University of Reading Malaysia campus, where she was Head of the Psychology Section. She has held various teaching and learning leadership roles on both campuses, and has a specific interest in assessment and large group teaching.

Dr Kai Li Chung is an associate professor in forensic psychology and individual differences at the University of Nottingham Malaysia. She is the Head of the School of Psychology and was previously the Head of Section and Programme Lead at the University of Reading Malaysia. She has a specific interest in assessment and cross-campus collaboration.

Lynnette Gnagi and Ainslie Chuah were senior student partners on this project, co-developing the focus group schedule and managing data collection across campuses.

Appendix A. TESTA categorisation of assessments pre- and postassessment reduction

Comparison of programme assessments at UoX/Y in 2018/19 to 2020/21, using Jessop and Tomas (2017) categorisations.

TESTA, Transforming the Experience of Students Through Assessment.

Appendix B: Focus Group Schedule

Characteristic	2018/19	2020/21
Number of summative assessments	61 High	43 Medium
Number of formative assessments	8 Medium	6 Medium
Percentage of marks from examinations	21 Medium	19 Medium
Variety of assessment methods	20 High	15 Medium

- 1. Introduction of focus group moderators and participants
- 2. Year 3 assessment structure
 - (a) Year 2 students (both campuses)
 - (i) This year all Part 3 optional modules at UoRM had 100% coursework assessment, whereas UoR modules were 25% coursework and 75% 'exam' (alternative assessment). How do you feel about these two-assessment structures? Which would you prefer to do, and why?
 - (ii) Due to Covid-19, so far you have only done online exams. How do you / would you feel about doing traditional exams for the first time in Part 3?
 - (iii) How are you planning to manage your workload in Part 3? Are you worried about anything?
 - (iv) What would be your ideal Part 3 optional module assessment structure? Should it be the same for all modules, or is it fine for the structure to differ?
 - (b) Year 3 students (both campuses)
 - (i) This year all UoRM Part 3 optional modules had 100% coursework assessment, whereas UoR modules were 25% coursework and 75% 'exam' (alternative assessment). How do you feel about this?
 - (ii) Do you feel you had sufficient support through formative assessment to complete the summative assessment?
 - (iii) Would any amendments improve this structure?
 - (iv) How did you find the balance in workload between your Project, compulsory modules, and optional modules?
 - (v) Is there anything you know now that you wish you'd known before?

- 3. Coursework and module selection
 - (i) Does the assessment structure of a module affect your decision-making process whilst selecting optional modules?
 - (ii) Does the type of assessment affect how you study and the learning approach you use?

Closing of session