

https://dspace.org/

Exploring the inner workings of the Southern African Geography Teachers Association Google group

Di Wilmont; Paul Goldschagg

https://hdl.handle.net/10539/39447
DHET

Downloaded from a DSpace repository. https://dspace.org/

Routledge
Taylor & Francis Group

Assessment & Evaluation in Higher Education

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

Unravelling an assessment puzzle: reflections on reconciling effective assessment and workload management in South African higher education

Naomi Nkealah

To cite this article: Naomi Nkealah (2023) Unravelling an assessment puzzle: reflections on reconciling effective assessment and workload management in South African higher education, Assessment & Evaluation in Higher Education, 48:8, 1092-1102, DOI: 10.1080/02602938.2023.2173139

To link to this article: https://doi.org/10.1080/02602938.2023.2173139

 Published online: 02 Feb 2023.

 Submit your article to this journal

 Article views: 202

 View related articles

 View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=caeh20

ASSESSMENT & EVALUATION IN HIGHER EDUCATION
2023, VOL. 48, NO. 8, 1092–1102
https://doi.org/10.1080/02602938.2023.2173139

Routledge
Taylor & Francis Group

Check for updates

Unravelling an assessment puzzle: reflections on reconciling effective assessment and workload management in South African higher education

Naomi Nkealah

University of the Witwatersrand, Johannesburg, South Africa

ABSTRACT

Assessment in higher education is a complex path to navigate, with unexpected twists and turns. Although many studies project the positive outcomes of assessment for learning, the outcomes do not always match the intention. Even when formative assessment is well-planned, the implementation may take such a toll, leaving a lecturer physically drained. A big puzzle with assessment in higher education is how to design formative assessment tasks that promote deep learning for students but do not increase staff workload. In this article, I explore this puzzle in relation to my teaching of a second-year module in an English course that is part of the Bachelor of Education programme at my university. I look at how the specific context of my teaching in 2021 influenced my formative assessment design and reflect critically on the affordances and constraints of the design in relation to existing conceptualisations of assessment in higher education. Lastly, I propose some possibilities for resolving the assessment puzzle and highlight the implications of these possibilities for the reconceptualisation of assessment in higher education.

KEYWORDS

Assessment puzzle; assessment in higher education; formative assessment; workload management

Introduction: the puzzle

In an article in the *City Press* newspaper on 23 July 2019, Fengu (2019) reports a case of the exclusion of a Wits University student, Oupa Segalwe, from the medical sciences programme owing to his attainment of a final examination mark of 59%, which was one mark short of the expected 60%. The assessment of Segalwe's examination had been done following the Cohen 60 method which was the standard method of assessment of medical sciences students at Wits. According to Fengu (2019, n.p.), "the Cohen method establishes an exam pass mark by using a fixed percentage of the marks of the best students". Taylor (2011, 678) explains that in this assessment method, "the pass mark is calculated as 60% of the score of the student at the 95th percentile". Based on this method, Segalwe was denied re-admission into his degree programme in 2019 because his final score was 59%, and not the expected 60%, after the calculations had been done. Segalwe and his family challenged the Cohen 60 assessment method which had disadvantaged him, to the extent that the case required the intervention of the Department of Higher Education, the Public Protector and the Human Rights Commission in South Africa.

CONTACT Naomi Nkealah naomi.nkealah@wits.ac.za
© 2023 Informa UK Limited, trading as Taylor & Francis Group

Comparing the Cohen method to norm-referenced assessment and the fixed Pass Mark (PM) assessment method, Taylor (2011, 678) notes that "the Cohen method provides more stable pass rates when compared to both norm-referenced (mean minus one standard deviation) and fixed PM (of 60%) methods". Although many universities adopt this method based on this understanding of its effectiveness in increasing pass rates, what emerges from Segalwe's story is a hidden flaw in the method, because it assumes that "the score of the student at the 95th percentile is an accurate indicator of exam difficulty and is consistent over time" (Taylor 2011, 678–679). Since Segalwe was unfortunate to find himself at the 59th position, he was excluded.

This case illustrates that assessment in higher education is a very delicate terrain. No matter how carefully one treads within it, the chances of something going wrong or someone experiencing disappointment are ever present. That an assessment method is carefully selected, thoughtfully planned and concisely executed does not mean the process will be favourable to everyone involved, including the assessor. The case of Segalwe is only one example to illustrate that there is no problem-free method of assessment; that assessment in higher education is only as good as it maximizes its affordances and mitigates its limitations.

For many university lecturers, choosing the type of assessment to implement in their courses – whether summative or formative or a combination of both – is not a big challenge; often the course structures and university systems dictate which type of assessment must happen when. Where lecturers may face a big challenge is in designing their particular type of assessment (e.g. formative assessment) to meet their course outcomes. As lecturers come in contact with students possessing different learning needs, they generally want to choose assessment methods that they deem best to enhance their students' learning. However, they may not be inclined to pursue demanding assessment methods that will jeopardise their own wellbeing. Consequently, a big puzzle with assessment in higher education is how lecturers can design formative assessment tasks that promote deep learning for their students but do not increase their teaching workload. The key question that this puzzle raises then is: are effective assessment and workload management reconcilable?

In this article, I explore this puzzle in relation to my teaching of a second-year module in the English II course which is part of the Languages, Literacies and Literatures curriculum in the Bachelor of Education programme at my university. I look at how the specific context of my teaching influenced my formative assessment design in 2021 and reflect critically on the affordances and constraints of the design in relation to existing conceptualisations of assessment in higher education. Lastly, I propose some possibilities for resolving my puzzle and highlight the implications of these possibilities for the reconceptualisation of assessment in higher education.

Conceptualising the puzzle

As a university lecturer, I am aware of the centrality of assessment in higher education course design. I personally value the role of formative assessment in helping my students improve their learning, thereby putting them in a position to perform well in summative assessments (Black and Wiliam 2009; Nicol 2009; Moeed 2015). I understand formative assessment as *assessment for learning* and summative assessment as *assessment of learning* and that the two work integratively to foster students' academic success (Moeed 2015; Black and Wiliam 2018). Many Wits University lecturers like me appreciate that the university Senate Policy on the Assessment of Student Learning (University of the Witwatersrand 2019) provides flexibility to allow for the adoption of different assessment methods across the university's multiple disciplines. Nevertheless, one must admit that the practice of assessment is fraught with challenges for both lecturers and students, and there is never a guarantee that one's chosen assessment methods will enable one to achieve their desired teaching and learning outcomes.

We cannot deny that sophisticated assessment increases lecturers' workload. In their article on authentic assessment, Villarroel et al. (2018, 841) note that "teachers are reluctant to change formal assessments, such as examinations, because changing these practices makes great demands on time, energy and intellectual resources". This reluctance is not only in relation to examinations; even changing formative assessments requires time, energy and intellectual resources. Biggs and Tang (2011) recommend constructive alignment in designing, teaching and assessing a course, such that the intended learning outcomes match the teaching and learning activities and these in turn match the assessment tasks. For this to be achieved, a lot of thinking has to go into the form and content, the scaffolding for students, the way in which one wants the tasks assessed (e.g. self-assessment, peer assessment or lecturer assessment), and the kind of feedback that one wants students to receive for each task (peer feedback, lecturer feedback or students' own self-evaluation). It is a great deal of planning, thinking and rethinking, deciding and un-deciding, making choices, reviewing, and finally accepting that the selected assessment tasks are fair, doable and ultimately geared at enhancing students' learning.

In the context of emergency remote teaching (ERT) in 2020–2021 owing to the global Covid-19 pandemic, thinking about assessment also involved thinking about whether one wanted students to engage with course tasks synchronously or asynchronously. All of this can be very time-consuming, energy-sapping and mentally exhausting, especially if one has multiple modules/courses they are teaching in one semester. Even after planning the assessment for a module/course, the implementation of it can turn out to be cumbersome, irrespective of whether it was designed as peer assessment, group assessment or self-assessment.

The literature on assessment repeatedly highlights the value of giving students feedback on their formative tasks so that they can improve on aspects they are not doing well, as well as be encouraged on what they are doing well (Black and Wiliam 2009; Nicol 2009; Moeed 2015; Black and Wiliam 2018). However, this assumes that students take time to read the feedback they receive. It also assumes that the feedback will be given timeously to allow students to apply it in their next assessment. Gibbs and Simpson (2004, 10) note that while lecturers spend an enormous amount of time providing comprehensive feedback to students, "feedback is often not read at all or not understood". Students tend to be interested more in marks earned than the feedback given. In the South African context, marks determine whether or not students will receive financial aid, which many students from previously marginalised groups (Black, Coloured and Indian) need to be able to study at university. So, while feedback on students' assessments is important, the amount of time lecturers spend on providing feedback needs to be weighed against the demands placed on them by large class sizes.

Many of the empirical studies on assessment in existing literature are undertaken in Europe, North America and Australia, where teaching assistance for staff is often available through graduate programmes. To illustrate, Nicol's (2009) study, in which he argues that regular self-assessment by students and regular teacher feedback enhances self-regulated learning for first-year students, was conducted at universities in the United Kingdom where the two courses he collected assessment data on had in-built capacity support for the lecturers, with graduate assistants running tutorial sessions and monitoring students' online activities while the lecturers focused on the delivery of teaching content. While this kind of support is not typical of UK higher education, Nicol's study does illustrate how some research on assessment tends to sensationalise the benefits of 'innovative' assessment methods while masking challenges on the ground that many lecturers face, such as huge class sizes, staff shortages and overwhelming marking loads. It is to this ugly side of assessment that this article speaks.

In the South African context where staff shortage is a persistent problem across universities and graduate assistanceship is not systemised, achieving the results Nicol presents in his work would be near impossible. With the massification of higher education, universities have become open to all school graduates who meet admission requirements and enrolment figures are rising each year (Fouche et al. 2021). At Wits University, we are increasingly finding ourselves teaching

very large classes, especially in first and second year courses in the Faculty of Humanities. As an example of the large student numbers in our courses, our English I course had 323 students in 2020, 320 in 2021, and 518 in 2022, according to the university's Learning Management System (LMS). So, if I designed formative tasks for my students where I plan on giving them individual feedback on each task to encourage deep learning (Biggs 1999; Moon 2004) and I have a class of over 200 students, then I am likely to find myself suffering from burnout and students experiencing frustration. I am either not going to meet up with the deadlines for the feedback or I am going to give superficial feedback that is not likely to help my students achieve the learning outcomes of my module. The big puzzle for me then is: how do I design formative assessment to enhance my students' learning and still be able to manage my workload and general wellbeing?

The puzzle context

I want to address the question above by undertaking a critical reflection of my formative assessment in a second-year Shakespeare module. The education context of my teaching means that lecturers are operating within a professional training framework such that graduates are expected to leave the university as qualified teachers of various school subjects. Thus, in the Wits School of Education (WSoE), we have content courses which focus on building subject content knowledge and methodology courses which centre on practice and the acquisition of skills. This is an important dimension of our context that has a bearing on how we design assessment in our courses: assessment design for any course would invariably depend on whether it is a content or a methodology course.

Within the Division of Languages, Literacies and Literatures specifically, I am involved in training students to become effective teachers of English who will end up teaching either English Home Language (English HL) or English First Additional Language (EFAL) at Senior Phase (Grades 7–9) as well as Further Education and Training (FET) Phase (Grades 10–12). The English II course is a content course that has three modules: 1) Grammar; 2) Shakespeare; and 3) Language, Literacy and Power. That the course is aimed at Senior Phase and FET Phase students is an important dimension that has consequences for the form our assessment takes, as assessment for an FET course cannot be designed the same as assessment for an Intermediate Phase course, for example. The English II course outline (University of the Witwatersrand 2021, 2–3) states that "our English undergraduate courses aim to develop the student-teacher's ability to… become a critical, independent thinker so that she/he, in turn, can assist her/his future learners to become critical independent thinkers". Assessment for this course must therefore be designed with this learning outcome in mind, and this requires careful planning.

In terms of class size, there were 201 students registered for English II in 2021, according to the course site on the university's LMS. That is a huge number of students to assess through both formative and summative assessments. The course runs for the whole year, with each module being taught in a teaching block. I taught the Shakespeare module in the second teaching block of 2021. The cohort of students in the course did not have the same attributes, as some of them had done English HL and therefore had studied Shakespeare's plays in schools while others (mostly from black township schools) had done EFAL and consequently had not encountered Shakespeare before now, unless they had done so in an out-of-school context. Students' school experience of assessment on Shakespeare was largely limited to contextual questions that tested only their knowledge of the content of texts, with little orientation toward the application of literary theories in interpreting literary texts. My assessment design therefore had to fill this gap by helping students to apply theory in reading the Shakespearean plays, while also assisting them to build knowledge about the content of the plays. Teaching Shakespeare in the context of decolonisation required that I design assessment tasks that both

recognise students' diverse backgrounds and positionalities and enable them to bring in their indigenous knowledges as they make meaning of the plays (Fouche et al. 2021).

Taking into consideration all these contextual aspects, it was inevitable that my formative tasks were going to have strengths and weaknesses as it takes a great deal of foresight to pre-empt problems that may arise in the long run and address them at the time of designing an assessment.

The assessment: reflecting on practice

In designing the formative assessment tasks for my Shakespeare module in 2021, my aim was to help students learn specific skills, such as literary reading and writing a critical literary essay, as these are fundamental skills for future teachers of English. This aim is in line with scholarly ideas on lifelong learning, disciplinary training, and constructively aligned teaching and learning (Gibbs and Simpson 2004; Nicol 2009; Biggs and Tang 2011). Because I teach literature and the Shakespeare module is part of a content course, it was important for me to see that my students can apply critical thinking in interpreting literary texts and to also draw on their own life experiences to make meaning out of texts, rather than just reproducing ideas abundantly available on the internet about Shakespeare's plays. The notion of reading literary texts through the medium of personal life experiences is in line with scholarship on literary reading as a socially responsive practice in which students' private experiences serve as powerful lenses for critical interpretation of texts (Allington 2012; Bomer 2017). Enhancing my students' socially responsive reading of Shakespearean plays was just as important to me as helping them to improve their academic performance in the module.

To teach Shakespeare's two plays, *Macbeth* and *A Midsummer Night's Dream*, in the context of ERT in 2021, I had to redesign my module to encompass both synchronous and asynchronous modes of teaching. Nicol (2009) notes that when redesigning a course, a course leader must have a clear understanding of what they want to achieve. Reflecting on this claim, I can say that I had a clear idea of what I wanted to achieve with the course redesign. First, I wanted to move away from the practice of using only multiple choice questions (MCQs) as my tool of formative assessment. In previous years when I taught the Shakespeare module, I used only MCQs as formative tasks. I did this as a way of sidestepping the challenge of marking over 200 essays within an often-limited timeframe. However, in 2021 I felt that MCQs alone did not push students to improve on their critical reading and writing which were learning outcomes for the English II course. Nicol (2009) notes that MCQs are very much about teacher control and less about student autonomy in learning. So, in moving away from an MCQs-only type of formative assessment, my goal essentially was to help my students to practise thinking and writing critically about texts rather than just knowing textual content. Yet, retaining some MCQs allowed me to set tasks that compelled students to read the texts, as reading the texts is a prerequisite for any critically engaging analysis of the texts that draw on theory and life experiences to conjecture meaning.

My second goal for redesigning the module to include weekly writing tasks was to get students into the habit of doing literary writing rather than just talking about texts as they did in the discussion forums that I set up on each play. Gibbs and Simpson (2004, 15) note that "probably the only way to gain facility with the discourse of a discipline is to undertake plenty of practice in using that discourse, for example through writing". Thus, I figured that in putting ideas down in writing, students would be more inclined to focus on the important points and also correct their language errors so that their English writing skills could improve. Writing tasks also help students to clarify their thinking about the ideologies embedded in literary texts and to learn to use correct terminology within the discipline to address these ideologies. In addition, the weekly writing tasks were meant to help students keep up the momentum that was built up during the weekly online tutorial sessions.

I had these clear goals in my mind when redesigning the module but my mistake was that I did not communicate these goals to the students as the scholarship recommends (Nicol 2009; Biggs and Tang 2011), neither during the first tutorial nor in the lecture schedule. Looking at my lecture schedule in retrospect and seeing that it had sections on module information, module description, text summaries, schedule of lectures and assessments but nothing on module goals makes me realise that the lecture schedule was missing a very important element. These module goals would have made very sound learning outcomes which Biggs and Tang (2011) insist are fundamental for the achievement of constructive alignment in teaching. As I think about my assessment practice retrospectively, I am able to see some of the faultlines which have contributed to my puzzle.

The redesign of the module itself brings me to the central issue in my puzzle: how to implement learning-enhancing formative assessment without increasing my workload. On the one hand, my workload was already overwhelming as my calculated annual hours were over 2500, which was above the university's 1800 hours maximum load for academics (University of the Witwatersrand 2020). Going beyond these hours could not have been good for my mental health and general wellbeing. Studies from South Africa and Australia have found that heavy workloads adversely affect staff wellbeing as academics with high teaching and supervision loads suffer from stress, anxiety, fatigue and professional burnout (Jasson, Du Plessis, and Simons 2022; Subban et al. 2022). These findings point to staff wellbeing as an important condition for staff performance and career enjoyment. Keeping my workload at a reasonably manageable level was an important consideration for my assessment design. On the other hand, I felt an obligation to give my students the best training I could possibly give, to prepare them for the classroom and for postgraduate studies.

I designed my assessment to have four weekly online writing tasks and two short MCQ tests. This is how the mark allocation was distributed:

Shakespeare Test 1 (*Macbeth*) – 30 marks
Shakespeare Test 2 (*A Midsummer Night's Dream*) – 30 marks
Shakespeare Writing Task 1 (*Macbeth*) – 10 marks
Shakespeare Writing Task 2 (*Macbeth*) – 10 marks
Shakespeare Writing Task 3 (*A Midsummer Night's Dream*) – 10 marks
Shakespeare Writing Task 4 (*A Midsummer Night's Dream*) – 10 marks

The marks for the six assessment tasks amounted to 100 marks, which made up the coursework marks for the module.

There were two MCQ tests of 30 marks each, with students answering 15 questions in each test (2 marks per question). The affordances of the MCQ tests, as noted earlier, were that they compelled students to read the plays, not just text summaries, and they tested for understanding of disciplinary knowledge such as figures of speech. The four writing tasks were to make up the rest of the marks (40%). Each writing task was to be 250–300 words only and students were expected to critically reflect on a given idea based on the lecture topic for that week. For example, Writing Task 4 was based on the topic of gendered representation in *A Midsummer Night's Dream* and it required students to compare the representation of one female character with that of one male character from a feminist perspective. Ample scaffolding was provided through online lectures and tutorials encouraging students to apply the theoretical ideas explained in lectures, to draw on their life experiences to make meaning of the texts, and to use textual evidence to support their arguments.

A criticism of my assessment design here is that while the multiplicity of tasks gave students opportunities to make up marks, it also overwhelmed students as they were busy with assessments almost every week. One disadvantage of assessment overload is that it leads to student dissatisfaction (Boud 2007), and yet I did not want my students to feel that way. Here lies another piece of the puzzle: how does one design formative assessment that keeps students engaged on a weekly basis but does not frustrate them?

The following assessment criteria for the writing tasks were prepared and given to students in advance:

- A concise argument or point. (2)
- Textual evidence of incident or character in play to illustrate the argument or point. (2)
- Relevant quotation of lines from play to buttress the point. (2)
- Appropriate connection of point to our everyday life experiences. (2)
- Logical, coherent and grammatically sound writing. (2)

Total marks = 10

This provision of assessment criteria was in line with the Biggs and Tang (2011) idea of assessment transparency and the Hattie and Timperley (2007) notion of feed up (as opposed to feedback), where you give students clear goals to which they should aim before they undertake an assessment.

However, students did not get an opportunity to engage with me and with each other about the criteria as the criteria were simply uploaded on the university's LMS, mainly because of time constraints. This is a weakness in my assessment planning as scholarship has shown that engaging students in assessment criteria clarifies the expectations and helps clear misunderstandings of the tasks (Biggs and Tang 2011; Shalem et al. 2013; Carless 2015). Matshedisho (2020, 169) states that "students may be disappointed with graded feedback when they realise that they misunderstood assessment task requirements". Thus, discussing assessment criteria with students gives them a sense of agency as they come to understand that they are important stakeholders in the design of the criteria.

Each writing task was to be completed during the week of the lecture topic: Task 1 for lecture week 2, Task 2 for lecture week 3, Task 3 for lecture week 5, and Task 4 for lecture week 6. This reflects the view that "assessment tasks distribute student effort evenly across topics and weeks" (Carless 2015, 63). There were seven weeks of lectures in total and lecture weeks 1, 4 and 7 did not have writing tasks. Lecture week 1 was the introductory lesson to Shakespeare and I felt that students needed to be immersed in the module before they could start writing about the texts. Lecture weeks 4 and 7 were on decolonising Shakespeare and that topic was reserved for the summative assessment in June. Although formative assessment is ideally an assessment that does not carry marks (Luckett and Sutherland 2000; Black and Wiliam 2018), I felt the need to attach marks to these tasks because our students are very much marks-driven, and in my experience tasks that do not carry marks are usually ignored. Gibbs and Simpson (2004, 8) note that "students will rarely write unassessed essays". The only way I thought I could motivate them to write was to make the writing tasks gradable. Carless (2015) notes that formative tasks to which marks are attached validate the exercise for students and give them an incentive to complete the tasks.

The problem with making the writing tasks gradable, however, was that (1) it massively increased my marking workload and (2) I was not able to meet up with the weekly marking of the tasks. The table below summarizes the marking that was involved and the number of hours.

As shown in Table 1, the marking hours were enormous and because of time constraints my feedback never reached students in time for them to apply it in their next task. Students received feedback for Task 1 after they had already submitted Task 2, and feedback for Task 2 after submitting Task 3, while their feedback for Tasks 3 and 4 reached them just before the examinations commenced in June 2021. Towards the examination in June, I had to ask for help from the division and two sessional staff were given to me, and together we were able to complete the marking of Tasks 3 and 4 just before the students wrote their examinations.

The use of marking assistants came with its own challenges of adding to the administrative workload, a problem highlighted by Fouche et al. (2021), as I had to have meetings with the

Table 1. Assessment marking load.

Writing tasks	Marking time	Number of students	Marking hours per week
Writing task 1	5 minutes	202	16.75
Writing task 2	5 minutes	202	16.75
Writing task 3	5 minutes	202	16.75
Writing task 4	5 minutes	202	16.75
Total hours = 16.75 hours x 4 weeks = 67 hours			

markers to brief them on the marking and discuss their marking to ensure that everyone was following the assessment criteria consistently, as well as moderating the marking. This is really the heart of my puzzle: I had good intentions in redesigning my module to include frequent formative tasks that promote deep learning for students, but in the end I failed to give students timeous feedback and I increased my workload tremendously. Here was a clear case of assessment boomerang where I personally suffered burnout in spite of my good intentions. So how could I have done this differently to ensure that students got timeous feedback and that my marking load did not quadruple?

Conceptualising possibilities for resolving the puzzle

One possibility for addressing my feedback dilemma would have been to use what Nicol (2009) calls model answers. In this case, I could have selected two or three of the best responses to each task and put them online on the course site for students to see what they ought to have done in their writing tasks, rather than trying to give individual feedback for each task written by each student. This could have saved me an enormous amount of time in the marking. However, model answers make feedback de-personalised whereas students want personalised feedback because it serves an affective purpose: it makes them feel good about their own work and motivates them to do better (Dawson et al. 2019). Moreover, using model answers to give students feedback on tasks was only going to solve the feedback aspect of my problem; it was not going to exempt me from marking 202 tasks each week. Since the tasks were set up to be graded, I had to spend time grading each one; whether I gave detailed feedback on each student's task or simply gave a grade and then gave generalised feedback to the class did not make a difference.

To solve the problem of the marking, which required plenty of time and needed to be done timeously, a possible solution could have been the use of peer assessment, which many scholars advocate as an effective strategy for enhancing students' evaluative judgement, self-regulated learning, reflective skills, task management skills and individual agency (Luckett and Sutherland 2000; Nicol 2009; Black and Wiliam 2018). Our university's LMS has a tool that can be used to randomly assign each student's task to another student to assess; it also allows for anonymous grading so as to prevent biased marking. Since students had the assessment criteria, it was possible for them to assess each other's work following the criteria, which was concise. However, many challenges exist with this approach.

First, and most importantly for me, peer assessment of this nature raises ethical questions (Moeed 2015) because it conflicts with my personal belief that any assessment that carries marks should be marked only by the lecturer or team of teaching staff who are sufficiently knowledgeable and experienced to make sound evaluative judgements. While peer assessment may sound fantastic in theory, empirical findings have shown that it is flawed in practice as students may assess each other's work based on personal biases as well as race and gender prejudices (Thondhlana and Belluigi 2017).

Second, Luckett and Sutherland (2000) note that reliability issues are a great concern in assessment where different assessors are involved, as inconsistencies are likely to emerge. If students did the marking, I would still have had to spend considerable time doing moderation of the

marks to ensure fair and consistent marking, and this may have been more demanding than if I did the marking myself. Nicol's (2009) claim that effective formative assessment can be achieved without increasing staff workload is vexing, because it does not hold true in my practical experience. Similarly, the assertion by Gibbs and Simpson (2004, 8) that "the trick when designing assessment regimes is to generate engagement with learning tasks without generating piles of marking" is short-sighted, because if learning tasks are set up to be graded so as to generate course marks for students, then there is no way of sidestepping piles of marking. Gibbs and Simpson (2004) themselves acknowledge that coursework marks make a higher contribution towards students' final marks than examination marks. Thus, what is evident here is the unfeasibility of and contradictions in some of the proposals and claims made in assessment literature.

A third significant problem that would have arisen with using peer assessment for my formative tasks is the failure of students to actually do the marking, which creates frustration for both lecturer and students (Fouche et al. 2021). This could have been the case for a number of reasons: students' sheer lack of time or the more strategic use of their time which would not prioritise such marking; internet connectivity problems which could have hindered access to the LMS; data limitations which could have prevented them from spending time to undertake online marking; and lack of experience in making evaluative judgements (Luckett and Sutherland 2000). This would have increased my workload in other ways as I would have had to spend time following up on students who did not do the marking and still do moderation.

A final possibility for resolving my puzzle of implementing effective formative assessment in my module without increasing my workload would have been to make the weekly writing tasks portfolio tasks that built up. Portfolio tasks have been found to be effective forms of formative assessment that enable students to receive valuable feedback on their work and improve the quality in preparation for submission as a summative assessment (Luckett and Sutherland 2000; Carless 2015). They are effective as assessment for learning because they encourage student self-assessment, peer review, student reflection on tasks, the provision of dialogic feedback to students on work in progress, and student achievement of diverse skills and capabilities (Carless 2015; Fouche et al. 2021). In this light, portfolio tasks would have been effective for my module in terms of eliminating the marking of the weekly tasks as they would have been set up as tasks that do not carry marks and the online peer review system could have worked in getting students to give each other peer feedback. Peer feedback, unlike peer assessment, demands less intensity of moderation and therefore less time.

However, for portfolio tasks to have been adopted as a form of formative assessment in this module, the course structure would have had to exclude a coursework mark component so that only a final summative assessment was required. This would then have allowed students to undertake peer reviewing of each other's tasks since they would not have been awarding marks, and I could have simply provided holistic feedback to the class on each writing task during each weekly tutorial session. Nevertheless, eliminating the coursework mark component would have disadvantaged students since this mark contributed significantly to their final scores. Effectively, it seems that although the portfolio task option seems like the ultimate solution to assessing large classes while still maintaining a manageable workload, it has its own problems.

Conclusion: implications for higher education assessment

It is clear from the evidence I have presented in this article that my assessment design had its strengths, but challenges of implementation meant it achieved its aims only partially. While a central aim was to give students feedback that would help them improve on their literary reading and critical writing skills within the module, my inability to meet up fully with the weekly marking did not allow students this learning opportunity and consequently they were short-changed. Like Segalwe, whose story was presented at the beginning of this article, many of my students would have felt that my chosen assessment method did not favour them. The

story of Segalwe shows that there is no problem-proof assessment since many medical sciences students in Segalwe's class were enabled to pass because of the Cohen 60 assessment method but Segalwe was unfortunate to end up with a final mark of 59% after the fixed percentage of the marks of the best students had been calculated. My critical reflection on my assessment design in my Shakespeare module confirms that assessments often have a drawback, for it was impossible to execute my formative assessment effectively to enhance my students' learning without increasing my workload.

Each of the possible resolutions to my puzzle described in the previous section has its affordances and constraints. This goes to corroborate the view that assessment in higher education is like walking through a forest path that is anything but straight. No matter which direction one tries to take assessment, there is always some kind of problem when it comes to implementation. Sometimes the problem is only evident on close scrutiny, and may affect only a handful of students, but it is still a problem. Carless (2015, 47) puts it bluntly: "all assessments have strengths and weaknesses". It seems therefore that before assessment can be designed one must take time to understand all the variables that may have an impact on the success of the implementation of the assessment. These variables include class size, course assessment structure, marking capacity available and marking timeframe. Often, it is not possible to know these things beforehand, and sometimes we are expected to have assessments in place even before we start teaching. All of these constrain effective assessment design and workload management.

The seeming irreconcilability between effective assessment and workload management, as demonstrated in this article, calls for a reconceptualisation of assessment in higher education to include the concept of strategic assessment where lecturers navigate the challenging terrain of providing feedback to a large class of students by adopting surface feedback where it is deemed that students can do with just that, and utilising deep feedback where it is necessary for students to pass a summative task. Adapting the ideas of surface learning, deep learning and strategic learning by students, as elucidated in the work of Biggs (1999) and Moon (2004), I propose that lecturers revert to strategising their formative assessments to alternate between giving deep feedback and surface feedback where appropriate. Given that increased workloads for university academics are a global problem, it is important that we rethink what assessments we choose and how we provide feedback in our courses. By alternating deep and surface feedback in relation to the learning outcomes of a course, rather than providing deep feedback all the way through, lecturers may perhaps manage their workloads and maintain their sanity.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Naomi Nkealah http://orcid.org/0000-0001-7953-9663

References

Allington, D. 2012. "Private Experience, Textual Analysis, and Institutional Authority: The Discursive Practice of Critical Interpretation and Its Enactment in Literary Training." *Language and Literature: International Journal of Stylistics* 21 (2): 211–225. doi:10.1177/0963947011435864.

Biggs, J. 1999. "What the Student Does: Teaching for Enhanced Learning." *Higher Education Research & Development* 18 (1): 57–75. doi:10.1080/0729436990180105.

Biggs, J., and C. Tang. 2011. *Teaching for Quality Learning at University*. 4th ed. Berkshire: Open University Press.

Black, P., and D. Wiliam. 2009. "Developing the Theory of Formative Assessment." *Educational Assessment, Evaluation and Accountability* 21 (1): 5–31. doi:10.1007/s11092-008-9068-5.

Black, P., and D. Wiliam. 2018. "Classroom Assessment and Pedagogy." *Assessment in Education: Principles, Policy & Practice* 25 (6): 551–575. doi:10.1080/0969594X.2018.1441807.

Bomer, R. 2017. "What Would It Mean for English Language Arts to Become More Culturally Responsive and Sustaining?" *Voices from the Middle* 24 (3): 11–15.

Boud, D. 2007. "Reframing Assessment as If Learning Was Important." In *Rethinking Assessment for Higher Education: Learning for the Longer Term*, edited by D. Boud and N. Falchikov, 14–25. London: Routledge.

Carless, D. 2015. *Excellence in University Assessment: Learning from Award Winning Practice*. London: Routledge.

Dawson, P., M. Henderson, P. Mahoney, M. Phillips, T. Ryan, D. Boud, and E. Molloy. 2019. "What Makes for Effective Feedback: Staff and Student Perspectives." *Assessment & Evaluation in Higher Education* 44 (1): 25–36. doi:10.1080/02602938.2018.1467877.

Fengu, M. 2019. "Public Protector Probes Wits Marking Method." *City Press*, July 23. https://www.news24.com/citypress/news/public-protector-probes-wits-marking-method-20190720.

Fouche, I., G. Andrews, L. Dison, and M. Prozesky. 2021. "Pedagogical and Decolonial Affordances of Group Portfolio Assessments for Learning in South African Universities." *Critical Studies in Teaching and Learning* 9: 76–98. doi:10.14426/cristal.v9iSI.325.

Gibbs, S., and C. Simpson. 2004. "Conditions under Which Assessment Supports Student Learning." *Learning and Teaching in Higher Education* 1: 3–31.

Hattie, J., and H. Timperley. 2007. "The Power of Feedback." *Review of Educational Research* 77 (1): 81–112. doi:10.3102/003465430298487.

Jasson, F. L., M. Du Plessis, and A. Simons. 2022. "Wellness Needs of Academic Employees: A Case Study of a South African University." *Journal of Psychology in Africa* 32 (3): 232–237. doi:10.1080/14330237.2022.2031628.

Luckett, K., and L. Sutherland. 2000. "Assessment Practices That Improve Teaching and Learning." In *Improving Teaching and Learning in Higher Education*, edited by S. Makoni, 98–130. Johannesburg: Wits University Press.

Matshedisho, K. R. 2020. "Straddling Rows and Columns: Students' (Mis)Conceptions of an Assessment Rubric." *Assessment & Evaluation in Higher Education* 45 (2): 169–179. doi:10.1080/02602938.2019.1616671.

Moeed, A. 2015. "Theorizing Formative Assessment: Time for a Change in Thinking." *The Educational Forum* 79 (2): 180–189. doi:10.1080/00131725.2014.1002593.

Moon, J. 2004. *A Handbook of Reflective and Experiential Learning: Theory and Practice*. London: Routledge.

Nicol, D. 2009. "Assessment for Learner Self-Regulation: Enhancing Achievement in the First Year Using Learning Technologies." *Assessment & Evaluation in Higher Education* 34 (3): 335–352. doi:10.1080/02602930802255139.

Shalem, Y., L. Dison, T. Gennrich, and T. Nkambule. 2013. "I Don't Understand Everything Here… I'm Scared': First-Year Education Students' Encounters with Assessment." *South African Journal of Higher Education* 27 (5): 1081–1098.

Subban, P., S. Laletas, E. Creely, J. Southcott, and V. Fernandes. 2022. "Under the Sword of Damocles: Exploring the Well-Being of University Academics during a Crisis." *Frontiers in Education* 7: 1004286. doi:10.3389/feduc.2022.1004286.

Taylor, C. A. 2011. "Development of a Modified Cohen Method of Standard Setting." *Medical Teacher* 33 (12): e678–e682. doi:10.3109/0142159X.2011.611192.

Thondhlana, G., and D. Z. Belluigi. 2017. "Students' Reception of Peer Assessment of Group-Work Contributions: Problematics in Terms of Race and Gender Emerging from a South African Case Study." *Assessment & Evaluation in Higher Education* 42 (7): 1118–1131. doi:10.1080/02602938.2016.1235133.

University of the Witwatersrand. 2019. "Senate Policy on the Assessment of Student Learning." http://intranet.wits.ac.za/exec/academic/qapo/Resources/Wits%20Institutional%20Policies/SSO%20%20Assessment%20of%20Student%20Learning.pdf#search=senate%20standing%20orders%20on%20assessment%20of%20student%20learning.

University of the Witwatersrand. 2020. "Academic Workload Policy." https://intranet.wits.ac.za/exec/registrar/Policies/Academic%20Workload%20Policy%202020.pdf#search=academic%20workload%20policy

University of the Witwatersrand. 2021. "English II Course Outline." https://ulwazi.wits.ac.za/courses/5393/files/search?preview=656401&search_term=course+outline

Villarroel, V., S. Bloxham, D. Bruna, C. Bruna, and C. Herrera-Seda. 2018. "Authentic Assessment: Creating a Blueprint for Course Design." *Assessment & Evaluation in Higher Education* 43 (5): 840–854. doi:10.1080/02602938.2017.1412396.