Rogaten, J., Clow, D., Edwards, C., Gaved, M., Rienties, B. (2020). Are assessment practices well aligned over time? A big data exploration. In M. Bearman, P. Dawson, R. Ajjawi, J. Tai, D. Boud (Eds). **Re-imagining University Assessment in a Digital World** (Vol. 7, pp. 147-164). Dordrecht: Springer. (20% contribution).

Are assessment practices well aligned over time? A big data exploration

Abstract
In the last twenty years a range of approaches have been adopted to facilitate Assessment of Learning as well as Assessment for Learning. With the increased interest in measuring learning gains using assessment data, it is important to recognise the potential limitations of using grades as proxies for learning. If there is a lack of alignment in terms of grade descriptors between modules within a qualification, students might perform really well on one module, and may underperform in a module that has relatively "harsh" grading policies. Using principles of Big Data, we explored whether students' grade trajectories followed a consistent pattern over time based upon their abilities, efforts, and engagement in two distinct studies. In Study 1, we explored a relatively large dataset of 13,966 students using multi-level modelling, while in a more fine-grained Study 2 we focussed on the pathways of students choosing their first two modules in six large qualifications. The findings indicated substantial misalignments in how students progressed over time in 12 large qualifications in Study 1. In Study 2, our analyses provided further evidence that students' grades did not seem to be well aligned. In all qualifications we found a highly significant effect of change over time depending on the achievement group. Based upon these findings, we provide clear recommendations how institutions might use similar insights into big data, and how they may improve the longitudinal alignment of grading trajectories by using consistent grading policies.
Key words: Assessment; learning analytics; longitudinal analysis;

Introduction

Assessment is a key driver for learning (Bearman et al., 2016; Boud, 2017; Coates, 2016). In the last twenty years a range of approaches have been adopted to facilitate Assessment of Learning (Boud, 2017; Boud & Falchikov, 2006; Coates, 2016), as well as Assessment for Learning (Bearman et al., 2016; Carless, 2007; Carless, Salter, Yang, & Lam, 2011). With the introduction of the Teaching Excellence Framework (TEF) in the UK, there is increased interest in measuring learning gains (Johnson, 2015; McGrath, Guerin, Harte, Frearson, & Manville, 2015). The broad assumption of the TEF is that universities that provide students with excellent teaching and learning opportunities will lead to high learning gains and value added, which will be financially rewarded[1].

One approach that is currently developed across a number of Office for Students projects[2] is to use students' academic performance as a proxy for estimating learning gains. This approach capitalises on the large quantities of student data routinely gathered by every university and may provide preliminary data-driven big data comparisons between different subjects, or even across different universities. Over the years, researchers and practitioners have tested a range of measurement approaches aiming to capture relative improvements in student learning (e.g., Anderson, 2006; Hake, 1998)[3]. Furthermore, using students' academic performance as a measure of learning progress has other advantages; firstly, it is widely recognized as a common proxy for mastery and learning. Secondly, grades are relatively free from self-reported biases, and thirdly, using academic performance allows a direct comparison of research finding with the results from other studies (Bowman, 2010; Gonyea, 2005; Rogaten, Rienties, & Whitelock, 2017). If a qualification (i.e., a set of modules and courses that build towards a certificate or degree) is well designed and assignments are aligned according to well-defined grade descriptors and/or rubrics (Bell, Mladenovic, & Price, 2013; Dawson, 2017; O'donovan, Price, & Rust, 2004), it would be reasonable to assume that as the level of difficulty increases, the grading over time will be adjusted.

---

[1] Although the specific details of the proposed measurements and metrics to be used for learning gains still have to be determined by the UK Government, future government funding might become related to students' learning gains as part of the teaching excellence narrative (Ashwin, 2017; Johnson, 2015).

[2] https://www.officeforstudents.org.uk/advice-and-guidance/teaching/learning-gain/

[3] The most common way of assessing learning gains is through use of pre-post testing (e.g., Dimitrov & Rumrill Jr, 2003). Although pre-post testing is considered as a standard and favourable approach for assessing learning gains, it can be a costly process, especially when measured across different modules.

Although the use of grades as proxies for learning gains sounds attractive, it is important to recognise the potential limitations of using grades as proxies for learning (Boud, 2018; Rogaten et al., 2017). A number of factors might explain why learning gains in a qualification, and grades in particular, might go up or down over time. Within the TEF framework, an assumption is that as students develop knowledge and skills in a qualification, students will strengthen their abilities to interlink concepts, to master key skills, and to be able to solve increasingly complex problems (Higher Education Commission, 2016; Johnson, 2015; McGrath et al., 2015). However, as argued by a recent opinion piece by Boud (2018) without clear learning outcomes that are embedded in a framework of explicit standards, it might be potentially inappropriate to compare assessment grades across modules and over time. At the same time, with the Open University UK (OU) the practice of awarding marks relative to the standard expected of students at that stage of the qualification, rather than expecting marks to increase substantially over the course of the qualification, seems widespread.

If there is a lack of alignment in terms of grade descriptors between modules within a qualification, students might perform really well on one module, and may underperform in a module that has relatively harsh grading policies. It is well known from the several studies carried out in to the reliability of assessment (e.g., Meadows & Billington, 2005; Moxley & Eubanks, 2015), that there can be substantial disparities and inconsistencies between (and even within) human markers, and there is evidence to suggest this may be a particular problem in distance education (Rienties, Clow, et al., 2017; Rienties, Rogaten, et al., 2017). Furthermore, as argued by Boud (2018) grades of a module may not be indicators of each of its learning outcomes, and when grades are aggregated across different learning outcomes, it may become impossible to determine what the respective outcomes each represent.

A particular concern relevant for our big data study is that Boud (2018, p. 5) argued that "[p]ass marks are determined within a disciplinary culture in relation to the internal features of the course unit and its tasks; no calibration of pass marks in relation to agreed standards is typically undertaken across units. Fifty per cent, say, is assumed to have a universal shared meaning and referent across different subject matter and different levels". Indeed, if we find large variations across modules across a qualification, this may imply that we may need to look at the potential alignments or misalignments between assessments within and across modules within a qualification.

Therefore, we used principles of Big Data (Ferguson et al., 2016; Rienties & Toetenel, 2016) in an online distance learning context to explore whether students' grade trajectories followed a consistent pattern over time based upon their abilities, efforts, and engagement in two distinct studies. This chapter presents these studies both in terms of the assessment practices they reveal but also as an illustration of how Big Data can inform institutional assessment practices. While we acknowledge the vast body of research on assessment practice, as also highlighted in this book, relatively few studies have used principles of big data to explore whether students' grade trajectories follow logical patterns (or not), and whether individual student characteristics might mediate these relations. Rather than hypothesis testing, in this chapter we primarily use an explorative study in one specific big data context, namely the OU.

Our exploration of students' grade trajectories from a large number of online students over time may help researchers to reflect upon whether the efforts of those who design assessments actually led to consistent assessment practices over time, and in particular whether those who graded students on their work did so in a consistent, intertemporal manner, ideally across a range of qualifications and disciplines. We specifically chose to conduct our study at the largest university in Europe, the OU, because great care and attention is provided towards designing and implementing modules and qualifications (Rienties & Toetenel, 2016; Toetenel & Rienties, 2016), and extensive quality assurances and practices are in place given the complex, large scale of educational provision amongst thousands of teachers and instructors at a distance (Richardson, 2013; Richardson, Alden Rivers, & Whitelock, 2015).

In Study 1, we used a relatively large dataset of 13,966 students to explore grade trajectories over time using multi-level modelling, whereby we acknowledge a hierarchical structure of the dataset through nesting data within three levels: module level and its related characteristics (e.g., module structure, workload, complexity of assessments, alignment of assessments with previous and follow-up modules); students level with its related characteristics (e.g., ability, socio-demographics); qualification level and its related characteristics (e.g., composition and sequence of modules to obtain a qualification). Therefore, our first research question is: To what extent are grade trajectories of students over time consistently aligned from one module to another, and how are these grade trajectories influenced by students' characteristics and qualification pathways?

Afterwards, in the more fine-grained Study 2, we focussed on six large qualifications with the OU, where we analysed the first two modules undertaken by students. We wished to explore how the paths that "new" students (in terms of studying for the first time at the OU) were taking through their qualification affected their achievement in terms of final marks of those first two modules. As highlighted by a wealth of research in higher education, and first-year experience in particular (Hillstock & Havice, 2014; Rytkönen, Parpala, Lindblom-Ylänne, Virtanen, & Postareff, 2012; Yorke & Longden, 2008), the transition in the first two modules is of essential importance for successful progression and continuation of study. Furthermore, recent large-scale research on students' experiences found substantial differences in how 16,670 "new" students experienced studying at the OU in comparison to 99,976 students who already completed several modules at the OU (Li, Marsh, Rienties, & Whitelock, 2017). If qualifications and introductory modules are well structured and assessment well-aligned, we would expect new students who are high achievers on their first module to tend to be high achievers on their second module, and low achievers on that module to be low achievers on the next (Conijn, Snijders, Kleingeld, & Matzat, 2017; Koester, Grom, & McKay, 2016; Popov & Bernhardt, 2013). Therefore, our second research question is: To what extent do the paths students take through the first two modules of their qualification impact their achievement in terms of marks?

Methods

Setting

Beyond our open-entry policy, another particular feature of the OU is that students may follow specific pathways within a qualification, or select modules and sequences of modules based upon their preferences. By mixing and matching modules over time, within certain qualifications students have substantial freedom to follow their interests, and select modules that fit with their interests, while other qualifications follow a more structured, fixed pathways. As previous research has found substantial differences between postgraduate and undergraduate learning designs (Li et al., 2017; Rienties & Toetenel, 2016), this study included only undergraduate modules that have run from 2013 to 2017[4].

Participants

For Study 1 a total of 13,966 students were included in a multi-level longitudinal analysis, whereby we selected several large qualifications for each of the four Faculties[5]. Students in this

---

[4] OU processes changed significantly at 2013, meaning comparison with studies undertaken prior to this time are difficult to make and of less interest to academics seeking analysis of current offerings.

[5] In terms of demographics, commonly more female students (61%) than male students (39%) study at the OU. Most students are from the UK (96%) from a white background. Students vary considerably in age, with 24%

sample all have achieved minimum grade of “pass” on all modules they were enrolled in. As such, this sample represents students who were continuously “successful”. For example, for students who were enrolled from October 2013 onwards, this would in practice mean that they would have needed to pass 4-6 modules consecutively to be included, while students who were enrolled from October 2015 onwards they would only need to have passed at least two modules. This is a very important caveat, as OU research and practice has consistently shown that many students are not always successful in terms of completing consecutive modules (Calvert, 2014; Li, Marsh, & Rienties, 2016; Li et al., 2017). One would assume that the selected student cohort (who passed all modules they were enrolled in) would continuously do well and perform with similar grades over time.

In Study 2, we selected a sample of students who passed at least two modules in the period from 2013 to the end of the 2016 calendar year. For each of the four Faculties, the top two qualifications in terms of student numbers were selected, apart from Faculty B, where only the top qualification was chosen, because the second most-popular qualification was fairly similar to the first. Within each qualification, the most popular paths taken by students were selected. As a result, 6794 students across these six qualifications were included in the analyses.

---

under 25 years old, 37% aged 26-35, 22% aged 36-45, 13% aged 45-55, and 5% aged 56 and over. More than half of students work full-time (53%), while 21% work part-time, 7% are looking after the home/family, and 5% are unemployed and looking for a job. Regarding learners’ qualifications, there are no formal academic entry requirements at undergraduate level at the OU. Around 40% of the students have A levels or equivalent (suggesting they had two or more years of post-compulsory schooling), 28% have less than A levels (suggesting they had not progressed beyond compulsory schooling), and around a quarter have a higher education or post-graduate degree. On average, 10% of the students report one or multiple disabilities. Participants in Study 1 and Study 2 were fairly similar in terms of demographics.

Data analysis

For Study 1, a 3-level growth curve model (Rasbash, Steele, Browne, & Goldstein, 2009; Rogaten et al., 2017) was fitted on student overall module grades taken each year starting from October 2013/2014 onwards[6] using MLWIN. In Study 2 the 6,974 students were split into three distinct achievement groups – high, medium, or low – based on their marks in their first module[7].

Results

Study 1

Figure 1 displays the qualification progression trajectories for some of the most popular qualifications at the OU. Each sequential module achievement is marked by the 'dot' on the

---

[6] Multilevel growth-curve modelling allows for estimating individual students' learning trajectories by fitting an overall average module curve and allowing each individual students' curve to depart from the average module curve. Using multilevel modelling it is possible to estimate what is the variance in students' initial achievements and their subsequent grade trajectories depending on what module they are enrolled in and whether students' initial achievements and grade trajectories depend on their individual differences and socio-demographic characteristics. As students at the OU can choose different pathways and elective modules (Edwards, 2017), not only can we compare how students progress within a qualification (e.g., Student 1 and Student 2) and which order of modules is most beneficial in terms of obtained grades, but we can also compare how students complete modules from other qualifications (e.g., Student 2 following Module 1 in Qualification 2; Student 3 following Module 2 in Qualification 3).

[7] I.e., Low = 40 – 59, Mid = 60 – 69, High = 70 +. For each qualification, a mixed ANOVA was carried out with time as a within-subject factor (first module to second module), and path (the top study paths, all others grouped) and achievement group (low, mid, high) as between-subject factors, and marks on the second module as the dependent variable.

line. Note that the straight lines between the assessment points are a result of regression modelling, whereby MLWIN predicts the best lines between assessment points. Of the twelve qualifications included in this analysis, all except one qualification (QUALC) had a negative trend over time, which indicates that students' performance in terms of grades dropped over time while going through a qualification. For example, the red line in Figure 1 represents QUALF, a science-related qualification, and there were six assessment points associated with the current progression of students in this qualification. On average, students in QUALF obtained 79.2 (SD = 11.3) for this first module, which is an above average grade. However, as indicated in Figure 1, the assessment scores over time deteriorated, with a final module score after 6 modules of 70.6 (SD = 13.2) for those students who passed all modules. In other words, "successful" students who completed all 6 modules in QUALF had a mean difference of -10.9 grade points.

➔ *Insert Figure 1 about here*

After students completed their first module, our modelling indicated that students were predicted to have high grades for the next module, with the average module grade of around 70. However, in contrast to our predictions the module grades dropped as students progressed from their first module to the next module (B = - 1.746), and the range of the drop between modules progression ranged between -4.566 and 1.074 grade points. For a detailed breakdown of the statistical analyses, see (Rienties, Rogaten, et al., 2017). As indicated in Table 1, the results showed that on average OU students performed relatively well in their first module (M = 70.8, SD = 4.7), but their grades dropped as students continued their studies towards qualification to an average of 68.8 (SD = 6.0).

➔ *Insert Table 1 about here*

Level 3 Role of qualifications on grade trajectories

Based upon the multi-level modelling the variance partition indicated that there was 12% of variance between the different qualifications (i.e., Level 3). Attainments achieved in any two qualifications correlated very weakly, and there was no strong alignment between selected undergraduate qualifications. In plain English, each qualification trajectory was quite independent from one another despite some groups of students taking the same modules for different qualifications. Of particular interest in the OU context, many students select and mix modules from different qualifications, so our findings might indicate that this might not be as beneficial (result in success) as perhaps hoped for. Importantly, some qualifications had high initial students' achievements on the first module, while others had relatively low initial achievements. In total, the qualification route accounted for 20% of variance in students' first module achievements. In other words, substantial differences are present when students start with a particular qualification, indicating a potential need to align introductory modules across the OU.

Furthermore, the particular qualification students were enrolled in determined their progress trajectories, and in total 30% of subsequent grade trajectories students made were due to being enrolled into a particular qualification degree. Qualifications that had first modules with relatively high average achievements tended to have a more rapid decline in their following semesters' average achievements, whereas qualifications with lower initial module grades had a lower decline. Note that given that nearly all qualifications had a negative grading trajectory, a lower decline in a way is an above average performance.

Level 2 Role of student characteristics in grade trajectories

The largest portion of variance in this model is at the student level (45%). The effort and time that students are able to put into studying will influence their performance within a module and across a qualification. Given the specific nature of OU students and the large impact of lifeevents on students' progression (Calvert, 2014; Li et al., 2017; van Ameijde, Weller, & Cross, 2016), it seems plausible that a large part of variation is explained by individual circumstances.

Given the widening access agenda of the OU (Li et al., 2017; Richardson, 2013; Richardson et al., 2015; van Ameijde et al., 2016) one would hope that students from a widening access background, who might initially struggle on the first module, will become more successful over time. However, our multi-level analyses indicated that students with below average achievements on their first module tended to have a steeper drop in their consequent module attainments. In contrast, students who obtained above average grades in their initial module had a lower drop in their subsequent module attainments.

Level 1 Impact of modules on grade trajectories

Lastly, 43% of variance lay on a module level, or "within-students", which indicated that there was a large proportion of variation (inconsistencies) between modules that form a particular qualification route. In plain English, if a student scored 70 on the first module and 70 on the second module, one would expect that this student would also score around 70 on the third module, fourth module, etc. However, substantial variation in module scores were present in all qualifications. As illustrated in Figure 2 of the QUALF, the actual scores of individual students across the six modules varied substantially.

➔ *Insert Figure 2 about here*

As illustrated by Figure 2 (1), the average score across the six modules in QUALF over time declined from module to module, with a notable exception for the fifth module. In terms of students' actual scores on their respective modules, Figure 2 (2) illustrates the wide variations in students' scores, whereby the lines from one module to the next show substantial variation. In other words, if modules would be "perfectly" aligned with a qualification (Boud, 2017;

Rienties, Rogaten, et al., 2017), one would expect flatter lines of individual student journeys, and variation would primarily be explained by individual ability, effort, and contribution.

Finally, Figure 2 (3) illustrates the predicted regression lines for each participant, which in most cases were downward sloping (i.e., indicating negative grade trajectories). Students who passed the first module with a good grade of 70+ were most likely to continue on follow-up modules, as more lines are visible in Figure 2 (3) for the third module onwards, relative to students with an initially low first module grade. In particular students with very high scores (80+) continued over time, and mostly had similar grades in subsequent modules. Students who score below 65 mostly performed worse for the second module, and were more likely to stop after the second or third module.

Study 2

In Study 2 we extended our analyses by looking specifically at the first two modules that "new" students took at the OU in order to determine whether we are providing a consistent practice at the start of their journey. As expected, many new students opted for a range of pathways after following their first module (e.g., QUALA, QUALF, QUALG)[8]. As highlighted previously in Table 1.2 in Rienties, Clow, et al. (2017), most students obtained a lower grade in their second module in comparison to their first module. In all but one instance, there was a significant time-path interaction ($p < .05$), and this was highly significant in five qualifications ($p < .001$). That is to say, with the exception of QUALE, students' grades changed over time depending on which study path they chose: some paths led to grades going up, and some to grades going down. Perhaps the most striking effect in this analysis is that there was a highly significant

[8] For example, in the two QUALA qualifications, all students started with module QUALAM1, but afterwards some students in the first QUALA selected QUALAM2, while others selected QUALAM3 or QUALAM4. Only for QUALD did all students in the sample follow the same second module.

time-achievement interaction in every single case ($p < .001$). That means that students in different achievement groups (high, mid, low) had different changes to their grades over time. If assessments were well aligned, we would expect achievement groups to be on average stable over time.

→Figure 3 about here

For example, Figure 3 shows an example of divergent changes in grades over time depending on path and achievement group for QUALA. Low achieving students on QUALAM1 (blue lines) tended to get markedly higher results on QUALAM2, but those who studied QUALAM4 got even higher grades – higher, in fact, than mid-achievers on QUALAM1 (green lines), who tended to decline in grades when they came to QUALAM4. High achieving students on QUALAM1 (red lines) achieved consistently high results when they got on to QUALAM2 (left-hand chart), but their grades declined if they instead choose QUALAM4 (right-hand chart). Again for a detailed statistical analyses of this and other qualifications, see Rienties, Clow, et al. (2017)

Implications for assessment practice

These studies indicate the potential value of Big Data to inform institutional assessment practices, which allowed us to track how students progressed over time in terms of grade trajectories, and how these were related to grading practices within and across modules within one distance learning institutions. After students completed their first module, our modelling indicated that students were predicted to have high grades for the next module, with the average module grade of around 70. However, in contrast to our predictions the module grades dropped as students progressed from their first module to the next module. As visually illustrated in Figure 1, this negative trend continued for most qualifications. A large part of this trend seems to result from institutional (module and qualification) factors.

Our multi-level analysis indicated that 12% of variance was explained on a qualification level, whereby each qualification trajectory was quite independent from one another despite some groups of students taking the same modules for different qualifications. Depending on the selected qualification, students' progression and grade trajectories in particular were significantly impacted. This is a relatively surprising finding, as in many qualifications students have substantial freedom to mix and match modules across the OU. Some qualifications seem to help students to obtain similar/comparable learning experiences and assessment outcomes, while more variation seems to be present in other qualifications. Our big data explorations highlight a potential need to better align expectations and modules within a qualification across the OU, as students get substantially different experiences depending on the respective qualification they are enrolled into.

The largest portion of variance in this model was explained by individual student characteristics (e.g., effort, ability, socio-demographics). Given the widening access agenda of the OU (Richardson, 2013; Richardson et al., 2015) one would hope that students from a widening access background, who might initially struggle on the first module, will become more successful over time. However, our multi-level analyses indicated that students with below average achievements on their first module tended to have a steeper drop in their consequent module attainments. In contrast, students who obtained above average grades in their initial module had a lower drop in their subsequent module attainments.

Lastly, another relatively surprising finding from Study 1 was that the students' journey from one module to another caused substantial transitional problems and imbalances in students' progression (43% of variance). Substantial variation in module scores were present while students were working through modules in all qualifications, which could be explained by inconsistent alignment of grade descriptors across a qualification and variations in marking within a module. This was further strengthened and confirmed in Study 2, where we saw a highly significant effect of the study path chosen on grades in the subsequent module. This was recently highlighted by Boud (2018), who noted that researchers need to tread carefully when comparing grades across time and discipline when the underlying frameworks of assessments and grading practices are not well aligned on an institutional level.

As discussed, there are many potential explanations for some of the particular instances observed: for instance, we would hope for some small improvement in grades for low-achieving students through our efforts to support them; alternatively, different modules teach and assess different skills and knowledge, for which aptitude may not be so strongly correlated. However, the size and consistency of the findings here suggest strongly that there are some serious discrepancies in assessment between modules on the same qualification. There are substantial challenges in aligning modules which have roles in multiple qualifications[9]. This adds extra weight to the recommendation to developing university-wide, cross-faculty processes for better aligning assessment and grading (Bearman et al., 2016; Boud, 2018; Dawson, 2017; Rienties, Clow, et al., 2017).

Considerations for practice

Based upon the findings from both studies, we identified three broad issues from our data: a) substantial freedom for students to select pathways; b) alignment of modules within a qualification; and finally c) alignment of marking across a qualification.

[9] For example, QUALAM3 appears in this analysis both as a second module for QAULA and as a first module for QUALB, which are located in different faculties (FacultyC and FacultyA)

Substantial freedom for students to select unique pathways: as highlighted by the detailed pathways that students can choose to complete a qualification, some programmes and qualifications have relatively fixed and structured pathways, whereby the options to choose different electives are limited. In contrast, other programmes and qualifications offer OU students wide and far reaching freedom to choose. However, like most other institutions the OU provides limited to no structural support about which pathways would fit students' needs and abilities, in contrast to other universities.

For example, in a large-scale adoption of Degree Compass, a course recommendation system, across two universities and two colleges in the US involving 40,000 students, Denley (2014) reported that the recommender analytics system steered students towards modules in which they were more likely to succeed. Similarly, in a large scale-adoption of E-advisor at Arizona State University, freshmen to sophomore retention rates increased from 76 to 84% (Phillips, 2013). Likewise, Denley (2014) found that a "six-year graduation rate… increased from 33 to 37.4%" (p. 65) when introducing course recommendations to students.

> Consideration 1: Institutions are encouraged to reflect on how to improve their communication to their students which modules fit with their needs and abilities, and be more explicit about successful pathways for students to obtain a qualification.

Furthermore, as highlighted by the recent Innovating Pedagogy Report (Ferguson et al., 2017), it is important that the institutions start to think about providing smart learning analytics to students and staff to help them to make the best decision of which qualification pathways might be the best way forward.

Consideration 2: Institutions may want to to explore whether (or not) to invest in smart learning analytics recommender systems that can help staff and students to support which paths within qualifications lead to highest success.

Alignment of modules within a qualification: As also highlighted by recent research at the OU (Nguyen, Rienties, & Toetenel, 2017; Nguyen, Rienties, Toetenel, Ferguson, & Whitelock, 2017; Rienties & Toetenel, 2016; Toetenel & Rienties, 2016) and these two studies in particular, students experience substantially different assessment practices and learning designs in general when transitioning from one module to another. Providing a consistent learning experience for students within and across a qualification will help students to adjust quickly and focus on their learning objectives, rather than spending a lot of time and effort trying to understand what is expected when a new module has a different design. Recent research in other institutions found similar inconsistencies in learning design and assessment practices (Bakharia et al., 2016; Mittelmeier et al., 2018). Therefore,

Consideration 3: Institutions should consider how they communicate and manage the students' expectations of the learning designs and assessment practices from one module to another.

Consideration 4: In the longer term, it would be beneficial to align the module designs across a qualification based upon evidence-based practice and what works, thereby allowing smooth transitions from one module to another in a qualification.

Alignment of marking within and across modules within and across qualifications: One potential explanation for large variations in grade trajectories in both studies is the effect of embedded expectations, norms and practice in relation to marking (Boud, 2017, 2018; Boud & Falchikov, 2006; Dawson, 2017). Across some qualifications there appeared to be a widespread deliberate approach of making early assessment relatively easy (Rienties, Clow, et al., 2017; Rienties, Rogaten, et al., 2017), both within modules (particularly the first assessment) and within qualifications (particularly the first module). This approach is intended to reduce drop-out, but may have unintended consequences. Furthermore, given that in most OU modules associate lecturers (external teachers hired to teach OU modules) are marking relatively small numbers of 10-20 assignments, potential misalignments might be present which may not be immediately apparent when just looking at average grades and the normal distribution curves currently used in quality assurance processes. Another potential explanation is that the increasing difficulty of the material being assessed may not be completely accounted for in the marks awarded. Final-year-equivalent modules rightly contain much more difficult material than entry modules. Ensuring that this is properly accounted for in marking expectations and practice is challenging, even if there was consensus that it should be.

> Consideration 5: It is good-practice that grades are aligned both within a module as well as across a qualification. For exam boards we recommend the inclusion of cross-checks of previous performance of students (e.g., correlation analyses) and longitudinal analyses of historical data to determine whether previously successful students were again successful, and whether they maintained a successful learning journey after a particular module.

Consideration 6: We recommend that clearer guidelines and grade descriptors across a qualification are developed, which are clearly communicated to staff and students to encourage effective uptake in the long-run.

Consideration 7: Given that many students follow modules from different qualifications, it is important to develop coherent university-wide grade descriptors and align marking across qualifications.

The use of Big Data reveals ways that assessment practices can be made more explicit that are otherwise "invisible" to educators and institutions. From these studies we found that even though substantial quality assurance and enhancement practices are in place in this institutional context, without a big data perspective the potential longitudinal misalignments of the complex students' journeys and assessment practices across several qualifications would not have been identified. By making these journeys and practices visible to staff, a start of a conversation can be made how to potentially improve the alignment of assessment practices over time.

Acknowledgements

This work is in part funded by the OfS grant "A longitudinal mixed-method study of learning gain: applying Affective-Behaviour-Cognition framework at three institutions". Study 1 was first published under a Creative Commons license as Chapter 1 in Rienties, B., Rogaten, J., Nguyen, Q., Edwards, C., Gaved, M., Holt, D., . . . Ullmann, T. (2017). *Scholarly insight Spring 2017: a Data wrangler perspective*. Milton Keynes: Open University UK. Study 2 was first published using Creative Commons as Chapter 1 in Rienties, B., Clow, D., Coughlan, T., Cross, S., Edwards, C., Gaved, M., . . . Ullmann, T. (2017). *Scholarly insight Autumn 2017: a Data wrangler perspective*. Milton Keynes: Open University.

References

Anderson, A. K. (2006). An Assessment of the Perception of Learning Gains of Freshmen Students in an Introductory Course in Nutrition and Food Science. *Journal of Food Science Education, 5*(2), 25-30. doi: 10.1111/j.1541-4329.2006.tb00078.x

Ashwin, P. W. H. (2017). Making sense of the Teaching Excellence Framework (TEF) results *Centre for Global Higher Education policy briefings* (Vol. 1). Lancaster.

Bakharia, A., Corrin, L., de Barba, P., Kennedy, G., Gašević, D., Mulder, R., . . . Lockyer, L. (2016). *A conceptual framework linking learning design with learning analytics.* Paper presented at the Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, Edinburgh.

Bearman, M., Dawson, P., Boud, D., Bennett, S., Hall, M., & Molloy, E. (2016). Support for assessment practice: developing the Assessment Design Decisions Framework. *Teaching in Higher Education, 21*(5), 545-556. doi: 10.1080/13562517.2016.1160217

Bell, A., Mladenovic, R., & Price, M. (2013). Students' perceptions of the usefulness of marking guides, grade descriptors and annotated exemplars. *Assessment & Evaluation in Higher Education, 38*(7), 769-788. doi: 10.1080/02602938.2012.714738

Boud, D. (2017). Standards-Based Assessment for an Era of Increasing Transparency. In D. Carless, S. Bridges, C. Chan, & R. Glofcheski (Eds.), *Scaling up Assessment for Learning in Higher Education. The Enabling Power of Assessment* (Vol. vol 5). Singapore: Springer.

Boud, D. (2018). Assessment could demonstrate learning gains, but what is required for it to do so? *Higher Education Pedagogies, 3*(1), 1-3. doi: 10.1080/23752696.2017.1413671

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education, 31*(4), 399-413. doi: 10.1080/02602930600679050

Bowman, N. A. (2010). Can 1st-Year College Students Accurately Report Their Learning and Development? *American Educational Research Journal, 47*(2), 466-496. doi: 10.3102/0002831209353595

Calvert, C. (2014). Developing a model and applications for probabilities of student success: a case study of predictive analytics. *Open Learning: The Journal of Open, Distance and e-Learning, 29*(2), 160-173. doi: 10.1080/02680513.2014.931805

Carless, D. (2007). Conceptualizing pre-emptive formative assessment. *Assessment in Education: Principles, Policy & Practice, 14*(2), 171-184. doi: 10.1080/09695940701478412

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education, 36*(4), 395-407. doi: 10.1080/03075071003642449

Coates, H. (2016). Assessing student learning outcomes internationally: insights and frontiers. *Assessment & Evaluation in Higher Education, 41*(5), 662-676. doi: 10.1080/02602938.2016.1160273

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS. *IEEE Transactions on Learning Technologies, 10*(1), 17-29. doi: 10.1109/TLT.2016.2616312

Dawson, P. (2017). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education, 42*(3), 347-360. doi: 10.1080/02602938.2015.1111294

Denley, T. (2014). How predictive analytics and choice architecture can improve student success. *Research & Practice in Assessment, 9*, 61-69.

Dimitrov, D. M., & Rumrill Jr, P. D. (2003). Pretest-posttest designs and measurement of change. *Work, 20*(2), 159-165.

Edwards, C. (2017). *Understanding student experience in the age of personalised study*. Paper presented at the ALT Online Winter Conference 2017. http://oro.open.ac.uk/53745/

Ferguson, R., Barzilai, S., Ben-Zvi, D., Chinn, C. A., Herodotou, C., Hod, Y., . . . Whitelock, D. (2017). *Innovating Pedagogy 2017: Open University Innovation Report 6*. Milton Keynes: The Open University.

Ferguson, R., Brasher, A., Cooper, A., Hillaire, G., Mittelmeier, J., Rienties, B., . . . Vuorikari, R. (2016). Research evidence of the use of learning analytics; implications for education policy. In R. Vuorikari & J. Castano-Munoz (Eds.), *A European Framework for Action on Learning Analytics* (pp. 1-152). Luxembourg: Joint Research Centre Science for Policy Report.

Gonyea, R. M. (2005). Self-reported data in institutional research: Review and recommendations. *New directions for institutional research, 127*, 73.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64-74. doi: 10.1119/1.18809

Higher Education Commission. (2016). From bricks to clicks. In X. Shacklock (Ed.), (pp. 1-76). London: Higher Education Commission.

Hillstock, L. G., & Havice, P. A. (2014). Exploring Characteristics of Retained First-Year Students Enrolled in Non-Proximal Distance Learning Programs. *Journal of College Student Retention: Research, Theory & Practice, 15*(4), 575-603. doi: 10.2190/CS.15.4.f

Johnson, J. (2015). *Fulfilling our potential: Teaching Excellence, Social Mobility and Student Choice*. (Cm9141). London: Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/474266/BIS-15-623-fulfilling-our-potential-teaching-excellence-social-mobility-and-student-choice-accessible.pdf.

Koester, B. P., Grom, G., & McKay, T. A. (2016). Patterns of Gendered Performance Difference in Introductory STEM Courses. *arXiv preprint arXiv:1608.07565*.

Li, N., Marsh, V., & Rienties, B. (2016). Modeling and managing learner satisfaction: use of learner feedback to enhance blended and online learning experience. *Decision Sciences Journal of Innovative Education, 14*(2), 216-242. doi: 10.1111/dsji.12096

Li, N., Marsh, V., Rienties, B., & Whitelock, D. (2017). Online learning experiences of new versus continuing learners: a large scale replication study. *Assessment & Evaluation in Higher Education, 42*(4), 657-672. doi: 10.1080/02602938.2016.1176989

McGrath, C. H., Guerin, B., Harte, E., Frearson, M., & Manville, C. (2015). Learning gain in higher education. Cambridge: HEFCE/RAND.

Meadows, M., & Billington, L. (2005). A review of the literature on marking reliability *Unpublished AQA report produced for the National Assessment Agency*.

Mittelmeier, J., Long, D., Melis Cin, F., Reedy, K., Gunter, A., Raghuram, P., & Rienties, B. (2018). Learning design in diverse institutional and cultural contexts: Suggestions from a participatory workshop with higher education leaders in Africa. *Open Learning*. doi: 10.1080/02680513.2018.1486185

Moxley, J. M., & Eubanks, D. (2015). On Keeping Score: Instructors' vs. Students' Rubric Ratings of 46,689 Essays. *WPA: Writing Program Administration, 39*, 53-80.

Nguyen, Q., Rienties, B., & Toetenel, L. (2017). *Unravelling the dynamics of instructional practice: A longitudinal study on learning design and VLE activities.* Paper presented at the Proceedings of the Seventh International Learning Analytics & Knowledge Conference, Vancouver, Canada.

Nguyen, Q., Rienties, B., Toetenel, L., Ferguson, F., & Whitelock, D. (2017). Examining the designs of computer-based assessment and its impact on student engagement, satisfaction, and pass rates. *Computers in Human Behavior, 76*(November 2017), 703-714. doi: 10.1016/j.chb.2017.03.028

O'donovan, B., Price, M., & Rust, C. (2004). Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teaching in Higher Education, 9*(3), 325-335. doi: 10.1080/1356251042000216642

Phillips, E. D. (2013). Improving Advising Using Technology and Data Analytics. *Change: The Magazine of Higher Learning, 45*(1), 48-55. doi: 10.1080/00091383.2013.749151

Popov, S. V., & Bernhardt, D. A. N. (2013). University competition, grading standards, and grade inflation. *Economic Inquiry, 51*(3), 1764-1778. doi: 10.1111/j.1465-7295.2012.00491.x

Rasbash, J., Steele, F., Browne, W. J., & Goldstein, H. (2009). *A User’s Guide to MLwiN*. University of Bristol: Centre for Multi-level Modelling.

Richardson, J. T. E. (2013). Approaches to studying across the adult life span: Evidence from distance education. *Learning and Individual Differences, 26*(0), 74-80. doi: 10.1016/j.lindif.2013.04.012

Richardson, J. T. E., Alden Rivers, B., & Whitelock, D. (2015). The role of feedback in the under-attainment of ethnic minority students: evidence from distance education. *Assessment & Evaluation in Higher Education, 40*(4), 557-573. doi: 10.1080/02602938.2014.938317

Rienties, B., Clow, D., Coughlan, T., Cross, S., Edwards, C., Gaved, M., . . . Ullmann, T. (2017). Scholarly insight Autumn 2017: a Data wrangler perspective. Milton Keynes: Open University.

Rienties, B., Rogaten, J., Nguyen, Q., Edwards, C., Gaved, M., Holt, D., . . . Ullmann, T. (2017). Scholarly insight Spring 2017: a Data wrangler perspective. Milton Keynes: Open University UK.

Rienties, B., & Toetenel, L. (2016). The impact of learning design on student behaviour, satisfaction and performance: a cross-institutional comparison across 151 modules. *Computers in Human Behavior, 60*, 333-341. doi: 10.1016/j.chb.2016.02.074

Rogaten, J., Rienties, B., & Whitelock, D. (2017). Assessing Learning Gains. In D. Joosten-ten Brinke & M. Laanpere (Eds.), *Technology Enhanced Assessment. TEA 2016. Communications in Computer and Information Science* (Vol. 653, pp. 117-132). Cham: Springer.

Rytkönen, H., Parpala, A., Lindblom-Ylänne, S., Virtanen, V., & Postareff, L. (2012). Factors affecting bioscience students' academic achievement. *Instructional Science, 40*(2), 241-256. doi: 10.1007/s11251-011-9176-3

Toetenel, L., & Rienties, B. (2016). Analysing 157 Learning Designs using Learning Analytic approaches as a means to evaluate the impact of pedagogical decision-making. *British Journal of Educational Technology, 47*(5), 981–992. doi: 10.1111/bjet.12423

van Ameijde, J., Weller, M., & Cross, S. (2016). Designing for Student Retention: The ICEBERG model and key design tips. In D. Whitelock (Ed.), *Quality Enhancement Series*. Milton Keynes: The Open University.

Yorke, M., & Longden, B. (2008). The first-year experience of higher education in the UK. York, UK: The Higher Education Academy,.

Figure 1 Qualification progression over time (estimated regression lines)

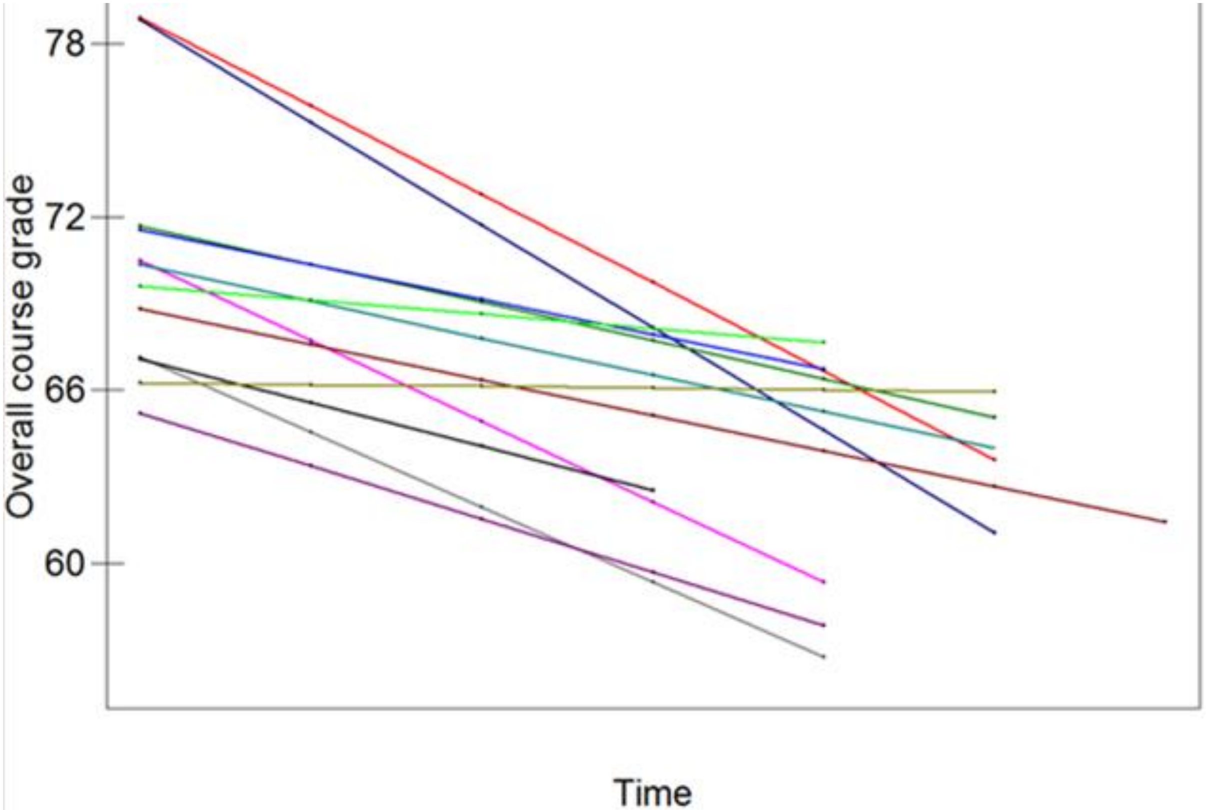
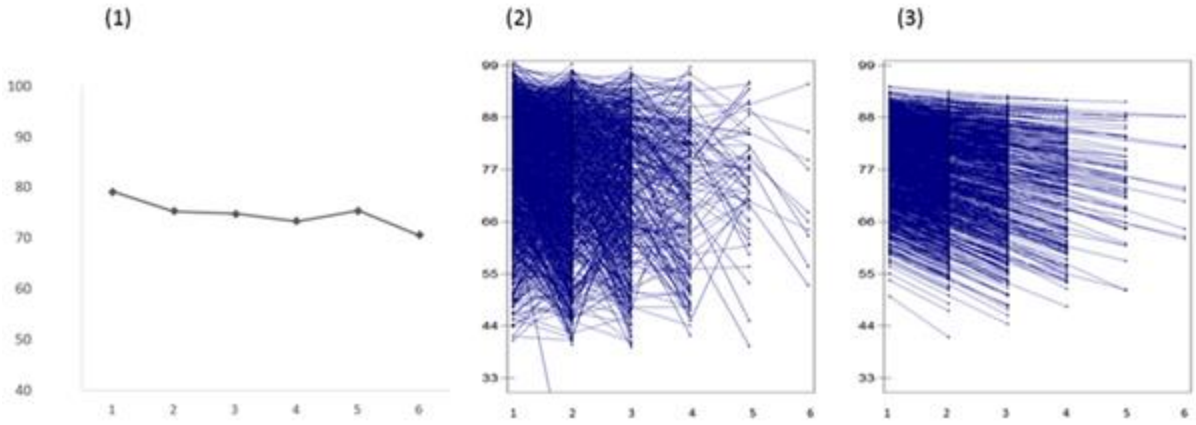


Figure 2 Module scores across six modules of QUALF



Note (1) Mean across modules in the qualification, (2) Actual module scores for each participant, and (3) trellis plot - predicted regression lines for each participant.

Figure 3: Mean grades for students on QAULA by achievement group, for those studying QUALAM1 then QUALAM2 (left-hand chart) and those studying QUALAM1 then QUALAM4 (right-hand chart).

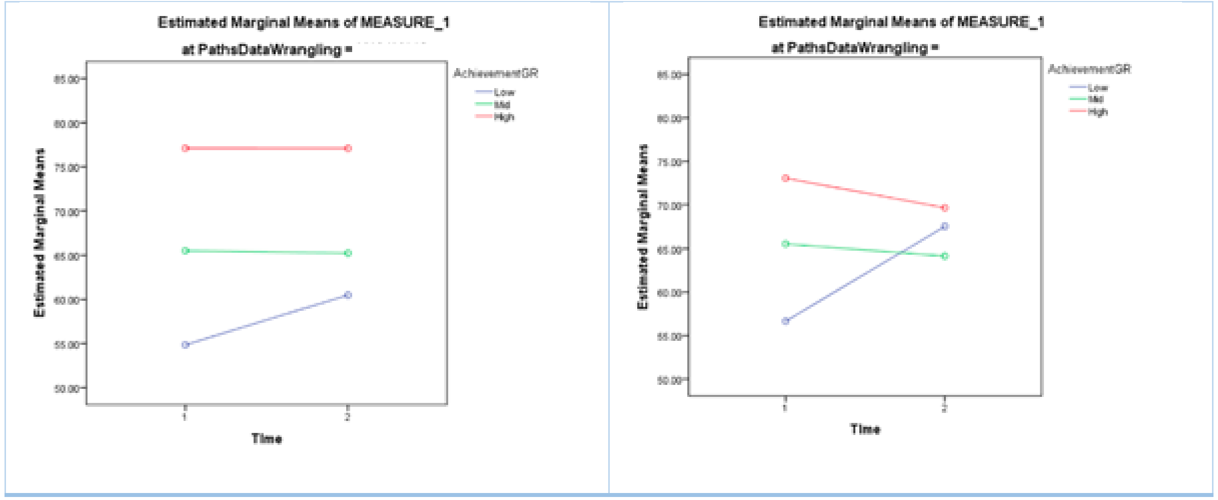


Table 1 Mean progression of students per qualification

	Qualification	Mean 1st module of all students who passed their first module (SD)	Mean last module of those who completed 5-6 modules (SD)
1	QUALF (n=1736)	79.2 (11.3)	70.6 (13.2)
2	QUALE (n=851)	80.3 (11.2)	73.9 (10.3)
3	QUALC (n=2629)	71.8 (10.1)	80.6 (10.8)
4	QUALC (n=405)	71.5 (11.4)	57.9 (9)
5	QUALD (n=968)	70.9 (9.4)	68.5 (11.8)
6	Open degree (n=3252)	70.4 (12.3)	69.9 (8.8)
7	QUALG (n=616)	69.9 (9.5)	68.5 (9.8)
8	QUALG (n=1079)	68.3 (10)	67.6 (9.4)

9	QUALA (n=599)	66.7 (10)	65.6 (9.0)
10	QUALC (n=423)	68.3 (10.1)	62.5 (2.1)
11	QUALG (n=344)	66.2 (9.1)	64.6 (12.9)
12	QUALG (n=980)	65.5 (10.5)	75 (10.0)
	Average across qualifications	**70.8 (4.7)**	**68.8 (6.0)**

Note: Qualifications were anonymised in six broad categories in line with OU Ethics policy. Multiple qualifications may be provided within each of these six broad categories.