JOURNAL OF CHEMICAL EDUCATION

pubs.acs.org/jchemeduc Article

Using Moodle Metrics to Analyze Student Navigation of Online Assessments with Mixed Question Types in Introductory Chemistry

Agnieszka Kosinska and Deborah L Gater*

Cite This: https://doi.org/10.1021/acs.jchemed.4c01249 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We present an analysis of the quiz metrics (question responses, marks, and times) recorded by a virtual learning platform (Moodle) in the context of a series of chemistry assessments. These metrics allow us to investigate whether any particular strategies are associated with higher or lower marks on the assessments. We find that there are no significant correlations between the order in which students attempt questions or edit their answers or between the percentage of the allowed time that students use and their performance on these quizzes. However, we did observe some patterns of behavior that seemed to distinguish students who obtained marks below the median and those who obtained higher marks. This work was conducted with reference to the model of "Scholarship of Teaching and Learning", and as such, we also describe some practical implications of this work for our own pedagogy, which others who teach chemistry across a range of educational levels may also find useful.

KEYWORDS: *Online Assessment, VLE, Chemical Education Research, Pedagogy, Sotl*

■ INTRODUCTION

The 2020 Covid pandemic resulted in a sudden and dramatic shift in the practicalities of both teaching and assessment in higher education (HE) in many institutions around the world. Much has already been discussed and published elsewhere on the (ever-evolving) impacts of these changes across the sector and within the chemical sciences specifically (for example, see chapters in ref 1). In our own subject, program and institution (chemistry on the Undergraduate Preparatory Certificate in Science and Engineering (UPCSE) at UCL), one significant alteration to our assessment was the extensive adoption of online tools for both coursework (asynchronous, noninvigilated) and test-type (synchronous, invigilated) summative assessment, facilitated via the institutional virtual learning environment (VLE).

During the immediate effort to translate assessments that had previously been predominantly issued on paper into suitable online formats, we shared many experiences with others in similar situations. For example, we had to consider the variety or type, number and content of questions,[2,3] as well as concerns relating to issues that could broadly be termed "academic integrity".[4,5] At that time, our general attitude might have been described as focused more on reducing or eliminating any perceived negative aspects of the enforced transition. However, once a "new normal" pattern of online assessment was established, sometime in 2021, we had an opportunity to reflect more deeply, using the ideas proposed by Mohamed et al.,[6] on the potential advantages or opportunities inherent in online assignments. The framework for this reflection broadly falls within the domains of the scholarship of teaching and learning outlined by Booth and Woollacott,[7] and was also informed by the principles for good practice in this field proposed by Felten.[8] In brief, the scholarship of teaching and learning approach seeks to encourage thoughtful professional development among educators as individuals but also in the context of a reflective community. Felten proposed 5 principles of good practice, which encompass the nature of the inquiry itself, being "focused on student learning", but also that such work should be "grounded in context", "methodologically sound", "conducted in partnership with students" and "appropriately public".[8] In a subsequent article, Booth and Woollacott further identified a series of internal domains where this practice takes place, ranging from the didactic and epistemic to the relational, ethical and ultimately societal context of pedagogy.[7] They also outlined four external contexts in which such scholarship takes place.

One opportunity arising from the transition to online assessment results from the existence of extensive and (relatively) accessible data that provide insight into student interactions with their assignments, that is not feasible with traditional paper-based scripts. Others have also reported their use of VLE logs to track factors affecting student outcomes in a range of HE courses, including (but not limited to) English as a foreign language (EFL) teacher training,[9] engineering exam attempts and retakes,[10] and computer networking.[11] Thus, the driving force for this study is 2-fold: first, to determine if there

Received: October 9, 2024
Revised: February 11, 2025
Accepted: February 12, 2025

Table 1. Details of Assessment Instruments Used to Obtain Data on Student Interactions with Moodle Quiz Assessments[a]

Instrument	Date	Content	Question number and marks per question	Question structure	Time-limit (min)	Format
CW1	mid-October (~1 week to complete)	atomic and electronic structure, periodic trends and mole concept	Q1 (4)	matching	90	asynchronous, open-book, one attempt.
			Q2 (4)	short answer		
			Q3 (12)	Cloze[b]		
			Q4 (8)	Cloze[b]		
			Q5 (1)	MCQ		
			Q6 (1)	MCQ		
			Q7 (1)	MCQ		
			Q8 (3)	calculation		
			Q9 (1)	MCQ		
			Q10 (5)	short answer		
HT1	mid-November (all students took quiz at same time in same room)	Lewis structures and VSEPR, titration, ideal gases, isotopes/mass spectra	Q1 (2)	Cloze[b]	45	in-person, invigilated in a computer room with screens not completely hidden from other students but invigilators walking around, closed-book.
			Q2 (8)	Cloze[b]		
			Q3 (11)	Cloze[b]		
			Q4 (6)	fill in blanks		
			Q5 (1)	MCQ		
CW7	December to January (over winter break)	kinetics	Q1 − Q8 (1 each)	MCQ × 8	90	asynchronous, open-book, one attempt.
			Q9 (5)	matching		
			Q10 (2)	calculation		
			Q11 (6)	fill in blanks		
			Q12 (10)	Cloze[b]		
			Q13 (10)	Cloze[b]		
HT2	late February (all students took quiz at same time in same room)	equilibria and buffers, kinetics	Q1 (2)	MCQ	45	in-person, invigilated in a computer room with screens not completely hidden from other students but invigilators walking around, closed-book.
			Q2 (3)	MCQ		
			Q3 (1)	MCQ		
			Q4 (3)	calculation		
			Q5 (9)	Cloze[b]		
			Q6 (9)	Cloze[b]		
CW16	late May to early June (~1 week to complete)	redox and electrochemistry	Q1 (6)	Cloze[b]	60	asynchronous, open-book, one attempt.
			Q2 (2)	short answer		
			Q3 (1)	MRQ		
			Q4 (1)	drag & drop image		
			Q5 (10)	Cloze[b]		
			Q6 (3)	Cloze[b]		
			Q7 (4)	Cloze[b]		

[a]**Abbreviations**: MCQ multiple choice question (one right answer only), MRQ multiple response question (multiple answers must be selected). [b]Cloze questions incorporate multiple question types within one scaffolded section. For examples, see Moodle community documentation.[12].

https://doi.org/10.1021/acs.jchemed.4c01249
J. Chem. Educ. XXXX, XXX, XXX−XXX

are any patterns in the way that students navigate online chemistry quizzes that are associated with either greater or lesser success in these assessment types; and second, to reflect on whether there are any implications for the way that practitioners either construct the quizzes or prepare students for online quizzes to improve the quality and reliability of these online assessment formats. These motivations give rise to the following specific research questions (RQs) that informed our data collection and analysis:

RQ1 How do students navigate online (Moodle) quizzes for summative and formative assessment in terms of: (a) the order in which questions are approached; (b) how often questions are revisited; (c) whether any question types—essay, structured (termed Cloze by Moodle), calculation, multiple choice questions (MCQs) etc—are more likely to be attempted "out of order" or to be revisited?

RQ2 Is there any correlation between the way that students navigate a quiz and their score in either the quiz as a whole or in a particular question subtype?

■ METHODOLOGY

Context

As previously stated, this study took place in the context of the 1-year intensive 'Undergraduate Preparatory Certificate in Science and Engineering' (UPCSE) course at UCL. This program is designed to prepare students for study in UK HE institutions. UPCSE is particularly aimed at students who have completed the equivalent of high-school outside the UK, but whose previous qualifications would not permit them direct admission to universities within the UK – particularly to Russell group institutions such as UCL itself. As such, all students enrolled in the chemistry subject had the majority of their prior education outside of the UK, and many were studying chemistry in English for the first time. All students applying for the course were obliged to take subject-specific entry tests in an online, multiple-choice format on Moodle. Beyond that, we have little definitive knowledge about their prior experience with online assessment. Due to the relatively small and diverse cohort (total enrollment in chemistry in the 22/23 academic year was 38 students), the decision was taken not to request or report any demographic information with respect to this study.

Students enrolled on the UPCSE program are expected to complete four subjects: English for Academic Purposes (EAP) and "Science and Society", which are compulsory for all students, as well as two subjects relevant to their future degree aim. Students taking chemistry may be additionally taking either biology or mathematics as their second subject. The chemistry syllabus and level are idiosyncratic to the UPCSE program. However, they are broadly aligned with the content of a UK A-Level in chemistry to prepare students for undergraduate studies in related fields. For those more familiar with the US educational system, much of the UPCSE chemistry syllabus overlaps with traditional General Chemistry courses, with additional Organic Chemistry elements.

There were four assessment categories in chemistry: final exam (including paper-based invigilated written and online oral components, in June); term-tests (paper-based, invigilated, in December and March); coursework (16 pieces of work for students to complete independently) and laboratory work (including a range of assessed pre- and postlab activities) that were assigned at regular intervals between October and May. The three half-term tests in October, February and May were counted toward the coursework grade. Individual pieces of coursework—including half-term tests—are intentionally "low-stakes" in terms of their contribution to the overall course grade, with coursework contributing 10% to the overall grade.

Ethics

Ethical approval for this project was sought from the Institution of Education at University College London and was awarded before data collection began. As DLG moved institutions after data collection was complete but before analysis was finished, ethical approval was additionally sought and awarded at Northeastern University London, and care was taken to follow suitable protocols with respect to data handling. The primary ethical concern in this project was that the researchers had dual roles as educators/assessors and as researchers. To mitigate this, care was taken in how information was provided and how consent was requested, and no data was downloaded or analyzed until after students had completed and left the program. All data was pseudonymized throughout analysis, and no demographic information was sought. The assessment instruments themselves retained a format used in previous years and followed the usual institutional quality control procedures for such instruments.

Assessment Instruments

Five assessment instruments were used as sources of data for this study: three online pieces of coursework (CW1, CW7 and CW16) that were largely similar to the previous year in format, delivery, content and in average grades (within one standard deviation of the two previous years) and two "half-term" tests (HT1 and HT2) for which new questions were written and for which format had alternated between online and paper-based in the two previous years, although the content and placement within the course remained the same as in previous years. Average grades in the HT tests varied slightly more than grades in coursework, given the variety of delivery modes (invigilated paper-based, invigilated online or noninvigilated online). In all cases, students could see the weightings of individual questions during the quiz itself, and were free to navigate between and edit responses to questions within the stated time-limit. Quizzes were organized to display a small number of questions on each page (depending on question length) and students had to click to proceed from one page to the next. The decision to split questions between pages was a compromise between limiting the number of page changes where a lag in loading might delay or frustrate a student within a time-limit, while also building in check-points where Moodle would automatically save answers on a page change, which could be retrieved even if a later technical issue affected submission. For each quiz, the type, weighting and order of questions seen by all students was the same, but many MCQ and matching questions had shuffled answers, calculation questions frequently gave each student a different value for a particular variable, and in some cases MCQ questions with similar topics and structures were drawn from a small bank of possibilities.

Note that CW1 was the first summative assessment attempted by students in chemistry, HT1 was the first invigilated test of any kind (paper or online), and CW16 was the last piece of coursework before the final exam period. Additional details of these tools are summarized in Table 1. Several question types (Matching, MCQ, MRQ, Drag & Drop) were automatically graded according to the mark scheme input into Canvas when the quiz was built. Some question types

https://doi.org/10.1021/acs.jchemed.4c01249
J. Chem. Educ. XXXX, XXX, XXX−XXX

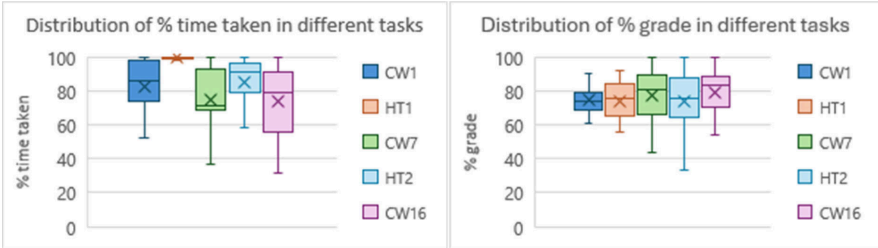

Figure 1. Distribution of marks and of time taken (both as percentages) for each Moodle assessment. Note × is the mean and the horizontal line is the inclusive median, with upper and lower quartiles bounded by the box. Whiskers show minimum and maximum values in each series.

(Short answer, Cloze, Calculation or Calculation with working) were either fully or partially manually graded by instructors according to a mark scheme. In some cases, manual grading involved updating the quiz autograding mark scheme to account for unforeseen entries by students, whereas in other cases the whole question was marked in the traditional manner with marks and feedback entered into the corresponding fields. These instruments represent all of the "Moodle quiz type" assessments completed by this cohort of students in chemistry during their course of study, but in each case, students completed other assessments within the category (e.g., coursework or half-term test) that were not administered as Moodle quizzes.

Data and Analysis

Data was downloaded from Moodle in two formats: (1) as standard Moodle quiz reports comprising .csv tables of attempts and outcomes, recording final answers and marks per question; and (2) as .pdf page prints of attempts, recording the time and nature of each answer or edit to an answer. The .csv data was converted to .xlsx format and relevant data regarding the times of each answer or edit to an answer from the .pdf files were also manually transferred to Excel. All statistical analysis was conducted within Excel. Analysis included descriptive statistics on grades and timings for quizzes and for individual questions, counts of numbers of edits and order of answers, tests for normality and ranked correlations. Tests for normality were conducted in order to determine whether a Pearson or Spearman correlation would be the most suitable measure in each case. All analysis was quantitative—there was no qualitative study of student responses.

■ RESULTS AND DISCUSSION

All students ($n = 38$) enrolled in chemistry in the 22/23 academic year were invited to participate in the project, and 22 students ultimately gave consent. This sample size gives a margin of error of ∼14% on any measured parameter, with 95% confidence. All of the participants attempted all five Moodle quiz assessments, with the exception of one student who failed to attempt CW16. One further student was awarded additional time to work on CW16 and their attempt has been excluded from analysis of timings.

A summary of the distribution of marks awarded and time taken for each assessment is provided in Figure 1. In general, the spread of times taken to complete the coursework (CW)-type assessments was greater than the spread of times taken in invigilated conditions (i.e., CW standard deviation values on times taken were 19%, 20% and 22% on CW1, 7 and 16, respectively, compared with standard deviations on times taken for half-term tests of 1% in HT1 and 14% in HT2). Similarly, students tended to use a smaller proportion of the allowed time to complete coursework compared with half-term tests: on average 82% of the allowed time for CW1, and 74% of the allowed time for both CW7 and 16 (despite the fact that the time limit for CW16 was shorter than the time limit for CW1 and 7, at 60 min rather than 90 min), compared with almost all of the allowed time in HT1 and 85% in HT2. Average marks across all five assessments were broadly similar at 74% for HT1, HT2 and CW1, 76% for CW7 and 79% for CW16. Standard deviations in marks ranged between 8% on CW1 to 18% on HT2. In general, most students, in most tasks, were able to choose to complete and submit their work before the automatic time-limit elapsed, and no single student took 100% of the allotted time for every task, indicating to us that the time-limits were generally appropriate for the work set.

Time Management

There were no consistent correlations between time taken on a particular instrument and mark obtained on the same instrument (Table 2) when comparing across all five quizzes.

Table 2. Results of Testing Spearman (Ranked) Correlation between Time Taken to Complete a Particular Assessment and Final Score Obtained on the Same Assessment

Assessment	Spearman correlation (time vs score)
CW1	−0.0164
HT1	not reported[a]
CW7	−0.1247
HT2	0.4477
CW16	−0.1650

[a]Results for HT1 are not reported because the narrow distribution of times taken in this case (see Figure 1) made any calculated correlation somewhat arbitrary.

The strongest correlation between time taken and mark was positive and occurred in HT2 (Spearman $r = 0.4477$, corresponding $R^2 = 0.20$). However, other assessments had weaker and negative correlations. The Spearman test was used to assess correlation in this case, as the data—particularly the distributions of time taken—were not consistently normal.

We then chose HT2 for further analysis of time spent on individual elements of the quiz, for two reasons. First the quiz had a relatively shorter time limit and was invigilated, so we anticipated that time-management would be a more significant factor in the students' efforts than in longer open-book assessments. Second, of the two shorter invigilated assessments (HT1 and HT2), there was a larger distribution of times taken in HT2. The full text of HT2 is available in Supporting Information.

https://doi.org/10.1021/acs.jchemed.4c01249
J. Chem. Educ. XXXX, XXX, XXX−XXX

Using the minute-by-minute time-stamps recorded by the VLE whenever a student moved from one question page to the next, we were able to examine in more detail the way in which students divided their time during HT2. However, this data must be interpreted with a degree of caution because it only records the times of page changes, and does not report what a student was actually doing. For example, it may be that a student worked out or reviewed one question that they had previously written down on paper while another question remained on the screen in front of them.

At the beginning of the attempts, there was, on average, a period of 3 min between a student starting their attempt and Moodle recording their first answer (range 1 to 6 min). At the end of the attempts, the average delay between the last recorded edit/answer and the submission was also 3 min, but here the range was larger—between 0 and 19 min. Only one student took the full allotted 45 min, and this student had no delay between their final edit/answer and the quiz being finished, suggesting that their attempt may have timed out and been submitted automatically.

In the case of MCQs there are some interesting features of the data when plotting the time spent working on the three MCQs (as a percentage of total allowed quiz time, including any editing) vs the marks achieved either for the MCQs themselves or for HT2 as a whole. On average, students spent 7 min working on MCQs (range 2 to 20 min). This corresponds to a percentage of the time allowed on the quiz of ~16% (mean) compared with the MCQ weighting of 22%. That is, students on average spent somewhat less time as a percentage of the allowed time on the questions than they were "worth" as a fraction of the marks. However, there is a negative correlation (Spearman −0.4047) between the time spent on MCQs and the overall quiz score, with two significant outliers who spent more than 30% of the total allowed quiz time on the MCQs but scored highly on the quiz overall. Removing the two outliers increases the magnitude of the correlation to −0.6694.

In comparison, students spent an average of 6 min (13% of allowed time, range 1 to 19 min) on the calculation question, Q4, compared with its grade value of 11% (see Figure 2 for the text of Q4). The grades awarded for Q4 were somewhat more binary, being either 0 for an incorrect answer (10 students) or 70−100% of the full question grade for a correct or partially correct answer (12 students). Students entering a numerically correct answer were awarded full marks, students who were wholly incorrect were automatically awarded zero marks and one student who was broadly numerically correct but had entered an answer with a rounding error was manually awarded partial credit with a comment relating to significant figures.

> A buffer solution is prepared by mixing 50.0 ml of a 0.500 M solution of sodium dihydrogen phosphate with 75.0 ml of a 0.28 M solution of phosphoric acid. Given that the pK_a of phosphoric acid is 2.14 and the Henderson Hasselbalch equation is:
>
> $$pH = pK_a + log_{10}\frac{[A^-]}{[HA]}$$
>
> What is the pH of the buffer solution?

Figure 2. Text of the calculation question 4 (Q4) as it appeared in HT2. Note that the concentration of the phosphoric acid solution was a variable in the question design, so each student saw a different value for that parameter. Note that the full text of HT2 is available in the Supporting Information.

There was no apparent correlation between time spent and grade achieved. However, the range of times used by students who provided an incorrect answer in Q4 generally appeared wider than the range of times used by students who answered this question correctly or mostly correctly (Figure 3). This may suggest a heterogeneity of approach in students who were faced with a question that they could not (or thought that they could not) answer—some simply did not take time to try, whereas others invested proportionally more time in their effort.

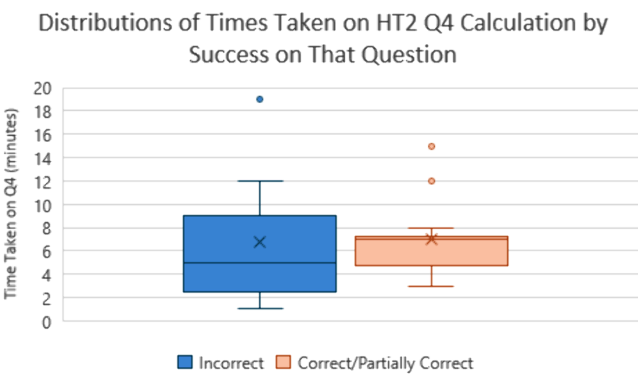

Figure 3. Distribution of time taken on the calculation Q4 in HT2 depending on whether the student obtained 0 credit (left/blue, n = 10) or some credit (right/orange, n = 12) on the question. Note × is the mean and the horizontal line is the inclusive median, with upper and lower quartiles bounded by the box. Whiskers show minimum and maximum values in each series, and outliers are shown as dots.

Question Order and Editing

We next examined the relationship between the order in which students attempted questions, again in HT2, and their associated marks. Of the 22 attempts at HT2, 8 students attempted questions in the order provided and 14 students did not. Similarly, we can see that 7 students edited their answers only after attempting all questions once, 14 students edited their answers as they went through the quiz and 1 student did not edit any of the answers that they entered. Of the 7 who answered questions in order and edited at least one answer, 5 students edited at the end of the quiz and 2 edited during the quiz. Of the 14 students who answered the questions out of order, the majority (n = 12) also edited during the quiz, and only 2 students answered the questions out of order but saved editing to the end.

Again, there was no significant difference in HT2 score depending on the strategy used, with average marks of 79% for those attempting questions in order vs 71% for those who did not, and average marks of 78% for those who edited only at the end vs 71% for those who edited as they went through (all averages within 1 standard deviation of the overall quiz average). Looking at these data from another perspective, 4 of the 11 students who scored below the class median answered the questions in order, and 4 of the 11 students who scored above the class median answered the questions in order. Similarly, 4 of the students who scored below the class median saved their editing to the end of the quiz, whereas 3 of the students who scored above the class median saved their editing to the end (and one did not edit at all).

We can also consider whether students were able to identify and edit incorrect answers to a question during the quiz itself, and we again analyzed responses to the calculation Q4 for this

https://doi.org/10.1021/acs.jchemed.4c01249
J. Chem. Educ. XXXX, XXX, XXX−XXX

purpose. In total, seven students entered an answer to Q4 and then edited their response subsequently. The remaining 15 students stayed with their original answers. The editing students included two who ultimately got full marks for the question, one who got partial credit and four who got 0, and those who did not edit included nine who obtained full credit and five who did got 0. Overall, there was a negative correlation (Spearman −0.3508) between the total number of edits that a student made across the whole quiz, and their total score on HT2. Together, these data are suggestive that many students were able to recognize within the quiz itself that they struggled with a particular question, and were able to choose to put in additional effort to "fix" an answer—whether or not they were successful.

Perspectives

In all cases—time taken, order of questions attempted and approach to editing—there was no significant relationship observed between a particular strategy and either higher or lower marks on the assessment. However, there were indications that the population of students who scored more poorly were in some cases more heterogeneous. They took a greater range of times to answer questions and they made more edits to their answers overall, and this may suggest they were potentially less strategic in their time management and quiz navigation. These observations are consistent with previous work that highlights the role of the affective domain,[13] self-efficacy[14] and the meta-cognitive aspects[15,16] of success in similar chemistry-based tasks. Overall, the picture is a complex one with multiple interacting factors affecting students feelings, motivations and perceptions. With respect to the affective domain, an extensive review by Flaherty et al.[13] emphasizes the importance of student feelings, and how the affective domain can affect academic performance. Similarly, with self-efficacy[14] and meta-cognitive aspects of learning[15,16] there is a recognized correlation between low self-efficacy or a lack of reflection and lower academic performance. However, the reasons for these apparent correlations are not yet fully understood and thus strategies that seek to invoke the affective domain to positively affect performance require additional research.[13]

In conversation with students, we might be encouraged to try to empower students to be aware of their use of time, and their ability to review and alter their answers during a quiz itself, as per suggestions regarding explicit teaching of metacognitive skills in ref 15. We might also be mindful of how we construct, disseminate and discuss practice or formative materials for students.[17] Although there is a tendency among educators and institutions to try to identify "at-risk" students[10,18] to target interventions, our own preference is to adapt some of the principles of 'Universal Design for Learning'[19,20] as we believe that attempts to improve learning and experience of assessment should benefit all.

As scholarly minded instructors, these observations encourage us to (continue to) use VLE data to analyze our assessment instruments. Such data permit fairly easy identification of what proportion of students is able to complete and review tasks, or whether tasks take undue amounts of time, and what fraction of students get each task correct. In addition to setting realistic time limits, we could also analyze assessments with respect to a content-related framework such as that suggested by Smith et al.,[21] to our own cognitive expectations,[22] or in light of studies examining perceived task difficulty.[23] Such reflection and analysis would allow us to construct quizzes and activities that are more authentically related to our pedagogical aims and practice.

We acknowledge that these perspectives must be interpreted cautiously due to the relatively small sample size and the limited study design of this work. A repetition of this study across larger cohorts may well illuminate factors that could not be identified here. Equally, the inclusion of demographic details and the use of broader research instruments such as surveys or interviews would allow for a more thorough understanding of these questions. One weakness of this work from a Scholarship of Teaching and Learning perspective is the lack of meaningful student partnership.[8] In part, this was an intentional act by the researchers given the otherwise complex power dynamics[24] involved in the project and the researchers' relative inexperience in this type of scholarship. Moving forward, we hope to share details of this project with current student cohorts to initiate new and more collaborative scholarship.

■ CONCLUSION

Analysis of VLE data that records student quiz attempts including time-stamps and edits to answers allowed us to answer—at least in part—our research questions as follows:

RQ1 Students took a range of approaches to navigating online quizzes in terms of the amount of the allowed time that they used, the proportion of time that they spent on different questions, the order in which they attempted questions and the amount and order of editing that they did.

RQ2 There was no evidence that the total amount of time spent on a particular quiz, or the order in which questions were attempted or edited, was correlated with overall score on that quiz. There was some indication that students who scored lower on a quiz might spend more time overall completing and editing multiple choice questions, and that the amount of time that these students spent on a calculation question was more heterogeneously distributed.

■ ASSOCIATED CONTENT

Data Availability Statement

For privacy reasons, individual pseudonymized data will not be shared. Other data and analysis can be shared on request.

Supporting Information

The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.4c01249.

Complete print-out of the preview version of the HT2 assessment instrument as it would appear to students during an attempt (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Deborah L Gater − *Department of Biosciences and Chemistry, CoMENS Faculty, Northeastern University London, E1W 1LP London, U.K.; Department of Chemistry and Chemical Biology, Northeastern University, Boston, Massachusetts 02115, United States;* orcid.org/0000-0001-5353-8179; Email: deborah.gater@nulondon.ac.uk

Author

Agnieszka Kosinska – *Centre for Languages and International Education, Institute of Education, University College London, WC1H 0AP London, U.K.*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jchemed.4c01249

Author Contributions

A.K. and D.G. contributed to project design, assessment instrument design and review, collection of consent, and paper writing. D.G. performed data acquisition and analysis and wrote the first draft of the paper.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This project received no formal funding. The authors would like to acknowledge their respective institutions for formal approvals and informal support of this work.

REFERENCES

(1) *Advances in Online Chemistry Education*; Pearsall, E., Mock, K., Morgan, M., Tucker, B. A., Eds.; American Chemical Society: Washington, DC, 2021; Vol. *1389*. DOI: 10.1021/bk-2021-1389.

(2) St. Hilaire, M. A.; Franco, J. Class and Exam Modalities' Impact on Student Exam Scores. *J. Chem. Educ.* **2023**, *100* (11), 4510–4513.

(3) Pilkington, L. I.; Hanif, M. An Account of Strategies and Innovations for Teaching Chemistry during the COVID-19 Pandemic. *Biochem. Mol. Biol. Educ.* **2021**, *49* (3), 320–322.

(4) Clark, T. M.; Turner, D. A.; Rostam, D. C. Evaluating and Improving Questions on an Unproctored Online General Chemistry Exam. *J. Chem. Educ.* **2022**, *99* (10), 3510–3521.

(5) DeKorver, B. K.; Krahulik, M.; Herrington, D. G. Differences in Chemistry Instructor Views of Assessment and Academic Integrity as Highlighted by the COVID Pandemic. *J. Chem. Educ.* **2023**, *100* (1), 91–101.

(6) Mohamed, M.; Rashid, R. A.; Alqaryouti, M. H. Conceptualizing the Complexity of Reflective Practice in Education. *Front Psychol* **2022**, *13*, 1008234 DOI: 10.3389/fpsyg.2022.1008234.

(7) Booth, S.; Woollacott, L. C. On the Constitution of SoTL: Its Domains and Contexts. *High Educ (Dordr)* **2018**, *75* (3), 537–551.

(8) Felten, P. Principles of Good Practice in SoTL. *Teaching & Learning Inquiry: The ISSOTL Journal* **2013**, *1* (1), 121–125.

(9) Liashuk, X. Relation between Moodle Activity and Student Performance in the Context of EFL Training in Higher Education. *J. Lang. Cult. Educ.* **2022**, *10* (1), 25–37.

(10) Kuzilek, J.; Zdrahal, Z.; Vaclavek, J.; Fuglik, V.; Skocilas, J.; Wolff, A. First-Year Engineering Students' Strategies for Taking Exams. *Int. J. Artif Intell Educ* **2023**, *33* (3), 583–608.

(11) Waspada, I.; Bahtiar, N.; Wibowo, A. Clustering Student Behavior Based on Quiz Activities on Moodle LMS to Discover the Relation with a Final Exam Score. *J. Phys. Conf Ser.* **2019**, *1217* (1), 012118.

(12) Embedded Answers (Cloze) question type. *Moodle.* https://docs.moodle.org/404/en/Embedded_Answers_(Cloze)_question_type (accessed 2024-06-18).

(13) Flaherty, A. A. A Review of Affective Chemistry Education Research and Its Implications for Future Research. *Chem. Educ. Res. Pract.* **2020**, *21* (3), 698.

(14) Willson-Conrad, A.; Grunert Kowalske, M. Using Self-Efficacy Beliefs to Understand How Students in a General Chemistry Course Approach the Exam Process. *Chem. Ed. Res. Pract.* **2018**, *19* (1), 265.

(15) Muteti, C. Z.; Zarraga, C.; Jacob, B. I.; Mwarumba, T. M.; Nkhata, D. B.; Mwavita, M.; Mohanty, S.; Mutambuki, J. M. I Realized What I Was Doing Was Not Working: The Influence of Explicit Teaching of Metacognition on Students' Study Strategies in a General Chemistry I Course†. *Chem. Ed. Res. Pract.* **2021**, *22* (1), 122.

(16) Chan, J. Y. K.; Bauer, C. F. Learning and Studying Strategies Used by General Chemistry Students with Different Affective Characteristics. *Chem. Ed. Res. Pract.* **2016**, *17* (4), 675.

(17) Knaus, K. J.; Murphy, K. L.; Holme, T. A. Designing Chemistry Practice Exams for Enhanced Benefits. An Instrument for Comparing Performance and Mental Effort Measures. *J. Chem. Educ.* **2009**, *86* (7), 827.

(18) Lewis, S. E.; Lewis, J. E. Predicting At-Risk Students in General Chemistry: Comparing Formal Thought to a General Achievement Measure. *Chem. Ed. Res. Pract.* **2007**, *8* (1), 32.

(19) Rose, D. Universal Design for Learning. *J. Spec. Educ. Technol.* **2000**, *15* (4), 47–51.

(20) Capp, M. J. The Effectiveness of Universal Design for Learning: A Meta-Analysis of Literature between 2013 and 2016. *Int. J. Inclusive Educ.* **2017**, *21* (8), 791–807.

(21) Smith, K. C.; Nakhleh, M. B.; Bretz, S. L. An Expanded Framework for Analyzing General Chemistry Exams. *Chem. Ed. Res. Pract.* **2010**, *11* (3), 147.

(22) Sanabria-Ríos, D.; Bretz, S. L. Investigating the Relationship between Faculty Cognitive Expectations about Learning Chemistry and the Construction of Exam Questions. *Chem. Ed. Res. Pract.* **2010**, *11* (3), 212.

(23) Schuessler, K.; Fischer, V.; Walpuski, M.; Leutner, D. The Moderating Role of Interest in the Relationship between Perceived Task Difficulty and Invested Mental Effort. *Educ Sci. (Basel)* **2024**, *14* (10), 1044.

(24) Acai, A.; Akesson, B.; Allen, M.; Chen, V.; Mathany, C.; McCollum, B.; Spencer, J.; Verwoord, R. E. M. Success in Student-Faculty/Staff SoTL Partnerships: Motivations, Challenges, Power, and Definitions. *Can. J. Scholarship Teach. Learn.* **2017**, *8* (2), 8 DOI: 10.5206/cjsotl-rcacea.2017.2.8.

https://doi.org/10.1021/acs.jchemed.4c01249
J. Chem. Educ. XXXX, XXX, XXX−XXX