EVALUATING DIFFERENT ASSESSMENT TYPES IN AN ONLINE GEOSCIENCE COURSE

Emre Dinç[1] and April L. Millet[2]
[1]*The Pennsylvania State University, University Park PA, United States of America*
[2]*The Pennsylvania State University, John A. Dutton e-Education Institute, University Park PA, United States of America*

ABSTRACT

This study investigated the effect of different assessment types on overall success in an online geoscience course covering energy's impact on the environment. Two groups took either the slower 1-credit or the faster 3-credit version of the course, which utilized the same content. Self-assessment questions developed using H5P, quizzes, summative assessment labs, and a capstone project were used to assess students. Independent samples t-test, multiple regression analysis, and repeated measures ANOVA were conducted for data analyses. Quiz success significantly differed between groups. Quiz, capstone, and lab assignment successes were significant positive predictors of overall success. The total frequency of answering H5P questions was a significant positive predictor of quiz and lab assignment successes. Frequencies for answering H5P questions and time spent on H5P questions in each module showed significant differences.

KEYWORDS

Assessment, Knowledge Checks, Online Course, Retrieval Practice, Geoscience Education

1. INTRODUCTION

The strategies need to be chosen to provide feedback and necessary support while designing online learning (Ally, 2008). However, identifying appropriate strategies to provide feedback in online learning was a challenge (Kearns, 2012). This was particularly observed during the COVID-19 pandemic with online assessment difficulties (Taufiq-Hail et al., 2021). Students engage with the assessment reflecting their learning during the online learning process more than engaging with each other (Şahin et al., 2020). Correspondingly, quizzes as a traditional assessment strategy, capstone projects as an assessment type connected to fieldwork and data collection in terms of active learning, and self-assessment resources as enabling metacognitive monitoring (Kearns, 2012; Motade & Deshpande, 2022; Yan & Brown, 2017) are used to evaluate learning and support the learning process. Ungraded self-assessment resources provide just-in-time and delayed feedback and engagement (Lewis et al., 2010; Wilkie et al., 2018), empowering students to oversee their learning (Nguyen et al., 2006). Quizzes also significantly affect the overall success and keep students focused on the course content (Salas-Morera et al., 2012).

Quizzes are often the first and essential assessment type instructors prefer to use in online courses before and during the COVID- 19 pandemic because quizzes in learning management systems encourage learners to engage with them (Rocco, 2007; Taufiq-Hail et al., 2021). Along with engagement, quizzes also affect student motivation positively (Raes et al., 2020). Weekly quizzes and self-assessment questions as active learning exercises create "highly structured courses" (Freeman et al., 2011, p. 176) and provide benefits to students getting prepared for learning (Hakk et al., 2011). Self-assessment, a metacognitive tool, is often used by successful students and positively affects overall success (Hartwig & Dunlosky, 2012). Capstone projects encourage personal meaning-making and contextual learning within the constructivist assessment strategies (Ally, 2008).

As observed during the COVID-19 pandemic, strategies with more qualitative approaches, such as capstone projects with checkpoints during the learning process, are more effective in online learning and online assessment (Montenegro-Rueda et al., 2021). Within the framework of engagement through quizzes

and lab assignments, self-assessing via H5P questions, and contextual learning with the capstone project, answers to these research questions (RQs) were sought:

1. Is there any significant difference in overall success and quiz, capstone, lab assignment achievements as well as the total frequency of answering H5P questions, total time spent on H5P questions, and the total points awarded for the correctly answered H5P questions between groups taking the course with different credits?

2. Is there any effect of quizzes, capstone, lab assignments, the total frequency of answering H5P questions, total time spent on H5P questions, the total points awarded on the correctly answered H5P questions, and weekly/biweekly duration modules on overall success?

3. Is there a significant effect of group (taking either 1-credit or 3-credit course) on the frequency of answering H5P questions when H5P questions are studied throughout the online course?

4. Is there a significant effect of group (taking either 1-credit or 3-credit course) on time spent on H5P questions when H5P questions are studied throughout the online course?

5. Is there a significant effect of the total frequency of answering H5P questions, total time spent on H5P questions, and the total points awarded on the correctly answered H5P questions on the quiz, capstone, and lab assignment achievements?

2. METHOD

2.1 Participants and Setting

A total of sixty-five participants took part in the study in the northeastern United States: eighteen participants from the 1-credit course and forty-seven participants from the 3-credit course. Both groups took the online course asynchronously with the same content throughout the five modules. Module1 – Why Energy Matters? describes the history of different types of resources used as energy. Module2 – What is Energy? focused on the basics of energy—how we talk about it, what it is, how much of it we use, and the relation between it and the US economy. Module 3 – Oil, Coal & Natural Gas | Drilling, Fracking & Reserves was about the formation and future of fossil fuels, the origins of oil, coal, and natural gas, and reserves and resources. Module 4 – Global Warming | Physics describes the physics associated with global warming. Module5 – Global Warming | History focused on the history of global warming specifically on who started the fire. Participants in the 1-credit course worked through one module per two weeks, whereas others in the 3-credit course worked through one module per week. 1-credit course was a part of a degree completion program.

2.2 Data Collection

Quizzes in each module were designed to check whether students mastered the science presented in the module. Students were given an opportunity to create an emission scenario for the next 200 years in the capstone project. Summative lab assignments focused on computer modeling for oil and climate in Module3 and Module4. H5P knowledge check questions were supplementary resources supporting learning the content throughout the modules. Points were assigned to correctly answered H5P questions; however, participants were informed that those points would not be aggregated into their overall course grade.

2.3 Data Analysis

Independent samples t-test was conducted to investigate the differences between groups taking the 1-credit and the 3-credit course. Multiple regression analysis was used to investigate the predictors for overall success as well as the quiz, capstone, and lab assignment achievements for all participants. Repeated measures ANOVA was conducted to investigate the differences between frequencies of answering H5P questions and time spent on H5P questions in each module.

3. RESULTS

3.1 RQ1. Independent Samples T-Test

There was a significant difference in quiz success ($t(62) = 2.464$, $p < 0.05$, Cohen's $d = 0.697$) between groups in favor of those taking the 1-credit course. Overall success ($t(62) = 0.637$, $p = 0.526$), capstone success ($t(62) = -0.611$, $p = 0.543$), lab assignment success ($t(62) = -0.594$, $p = 0.555$), the total time spent on H5P questions ($t(62) = 1.957$, $p = 0.055$), the frequency of answering H5P questions ($t(62) = 0.542$, $p = 0.590$), and the total points awarded for correctly answered H5P questions ($t(62) = -0.197$, $p = 0.844$) did *not* significantly differ between groups taking 1-credit and 3-credit.

3.2 RQ2. Multiple Regression Analysis for Overall Success

Multiple regression analysis was conducted for all students to investigate the effect of different durations per module and assessment types (Table 1). One outlier was detected based on Mahalanobis distance and removed from all analyses. Tolerance values were higher than 0.2, and VIF values were less than 10. The ANOVA result showed significance ($p < 0.001$), which is a good fit for the data. Different duration per module between groups ($p = 0.595$) did not significantly affect overall success. Quiz ($p < 0.001$), capstone ($p < 0.001$), and lab assignment ($p < 0.001$) achievements were significant predictors of overall success in the course. The unstandardized coefficient for quiz success is 0.469. Each one-unit increase in total quiz score is an increase in the overall score of 0.469 points. When we check their standardized coefficients, capstone project success ($B = 0.431$) was more powerful than quiz success ($B = 0.425$) and lab assignment success ($B = 0.341$) in predicting the overall success.

Table 1. Multiple regression analysis results for overall success

	Unstandardized coefficients		Standardized coefficient B	t	p
	B	Std. Error			
constant	10.746	2.890		3.719	< 0.001
Quiz success	0.469	0.036	0.425	12.941	< 0.001
Capstone success	0.214	0.012	0.431	17.376	< 0.001
Lab assignment success	0.699	0.071	0.341	9.857	< 0.001
Total frequency of H5P questions	-0.112	0.115	-0.047	-0.970	0.336
Total time spent on H5P questions	1.542e-5	0.000	0.017	0.761	0.450
Total points awarded for H5P questions	0.122	0.110	0.049	1.105	0.274
Groups	-0.398	0.746	-0.013	-0.534	0.595

Note. *Groups* variable is for taking either the 1-credit course or the 3-credit course.

3.3 RQ3. Repeated-Measures ANOVA for Frequencies

Since the sphericity was violated ($p < 0.001$), we interpreted the Greenhouse-Geisser correction ($\varepsilon = 0.687$). A repeated-measures ANOVA was conducted with group as the between-subject factor and frequencies of answering H5P questions in each module as the within-subject factor. The overall model was significant; Wilks Lambda = 0.283, F (4, 59) = 37.414, $p < 0.001$, partial eta squared = 0.717. There was no significant effect of group on frequencies, F (1, 62) = 0.293, $p = 0.590$, partial eta squared = 0.005, supporting the non-significant result of the first research question related to frequencies.

There was a significant effect of specific module on frequencies, F (3, 170) = 65.303, $p < 0.001$, partial eta squared = 0.513. Pair-wise comparisons with Bonferroni adjustment for the frequencies of answering H5P questions (Table 2) in each module for all participants indicated: The frequency of answering H5P questions in Module1 was significantly higher than frequencies in Module2 ($p < 0.001$, Cohen's $d = 1.646$), Module3 ($p < 0.001$, Cohen's $d = 0.876$), Module4 ($p < 0.001$, Cohen's $d = 0.968$), and Module 5 ($p < 0.001$, Cohen's $d = 1.530$). The frequency of answering H5P questions in Module3 was significantly higher than frequencies in Module2 ($p < 0.001$, Cohen's $d = 0.997$) and Module5 ($p < 0.001$, Cohen's $d = 0.863$). The frequency of answering H5P questions in Module4 was significantly higher than frequencies in Module2 ($p < 0.001$, Cohen's $d = 0.569$) and Module5 ($p < 0.001$, Cohen's $d = 0.504$).

Table 2. Pairwise comparisons of frequencies of answering H5P questions

Module	Module	Mean Difference	Std. Error	p	95% Confidence Interval for Difference	
					Lower	Upper
1	2	2.942*	0.257	< 0.001	2.194	3.680
	3	1.658*	0.258	< 0.001	0.907	2.409
	4	1.978*	0.256	< 0.001	1.234	2.722
	5	2.882*	0.276	< 0.001	2.080	3.684
2	1	-2.942*	0.257	< 0.001	-3.689	-2.194
	3	-1.283*	0.173	< 0.001	-1.788	-0.779
	4	-0.964*	0.170	< 0.001	-1.458	-0.469
	5	-0.059	0.142	1.000	-0.472	0.354
3	1	-1.658*	0.258	< 0.001	-2.409	-0.907
	2	1.283*	0.173	< 0.001	0.779	1.788
	4	0.320	0.169	0.631	-0.172	0.812
	5	1.224*	0.166	< 0.001	0.742	1.706
4	1	-1.978*	0.256	< 0.001	-2.722	-1.234
	2	0.964*	0.170	< 0.001	0.469	1.458
	3	-0.320	0.169	0.631	-0.812	0.172
	5	0.904*	0.172	< 0.001	0.404	1.404
5	1	-2.882*	0.276	< 0.001	-3.684	-2.080
	2	0.059	0.142	1.000	-0.354	0.472
	3	-1.224*	0.166	< 0.001	-1.706	-0.742
	4	-0.904*	0.172	< 0.001	-1.404	-0.404

3.4 RQ4. Repeated-Measures ANOVA for Time Spent

Since the sphericity was violated ($p < 0.001$), we interpreted the Greenhouse-Geisser correction ($\varepsilon = 0.635$). A repeated-measures ANOVA was conducted with group as the between-subject factor and time spent on H5P questions in each module as the within-subject factor. The overall model is significant; Wilks Lambda = 0.728, F (4, 59) = 5.498, $p < 0.001$, partial eta squared = 0.272. There was no significant effect of group on time spent, F (1, 62) = 2.011, $p = 0.161$, partial eta squared = 0.031, supporting the non-significant result of the first research question related to time spent.

There was a significant effect of specific module on time spent on H5P questions, F (3, 158) = 3.617, $p < 0.05$, partial eta squared = 0.055. Pair-wise comparisons with Bonferroni adjustment for the total time spent on H5P questions (Table 3) in each module for all participants indicated that time spent on H5P questions in Module4 was significantly higher than time spent on H5P questions in Module5 ($p < 0.05$, Cohen's $d = 0.458$). There were no other significant differences in time spent on H5P questions in modules.

Table 3. Pairwise comparisons of time spent on H5P questions

Module	Module	Mean Difference	Std. Error	p	95% Confidence Interval for difference	
					Lower	Upper
1	2	3122.225	1231.835	0.138	-463.607	6708.057
	3	1449.104	1380.341	1.000	-2569.022	5467.230
	4	364.565	1519.389	1.000	-4058.327	4787.457
	5	3147.002	1258.705	0.151	-517.048	6811.052
2	1	-3122.225	1231.835	0.138	-6708.057	463.607
	3	-1673.121	825.336	0.469	-4075.648	729.405
	4	-2757.660	992.777	0.072	-5647.600	132.280
	5	24.777	396.779	1.000	-1130.235	1179.788
3	1	-1449.104	1380.341	1.000	-5467.230	2569.022
	2	1673.121	825.336	0.469	-729.405	4075.648
	4	-1084.539	1288.430	1.000	-4835.115	2666.037
	5	1697.898	832.989	0.458	-726.905	4122.701
4	1	-364.565	1519.389	1.000	-4787.457	4058.327
	2	2757.660	992.777	0.072	-132.280	5647.600
	3	1084.539	1288.430	1.000	-2666.037	4835.115
	5	2782.437*	833.719	0.014	355.508	5209.366
5	1	-3147.002	1258.705	0.151	-6811.052	517.048
	2	-24.777	396.779	1.000	-1179.788	1130.235
	3	-1697.898	832.989	0.458	-4122.701	726.905
	4	-2782.437*	833.719	0.014	-5209.366	-355.508

3.5 RQ5. Multiple Regression Analysis for Quiz, Capstone, and Lab Achievements

Multiple regression analysis was conducted for all students to investigate the effect of the total frequency of answering H5P questions, total time spent on H5P questions, and the total points awarded on the correctly answered H5P questions on the quiz, capstone, and lab assignment achievements (Table 4). Tolerance values were higher than 0.2, and VIF values were less than 10. The ANOVA results showed significance ($p < 0.001$), which is a good fit for the data. The total frequency of answering H5P questions was a significant positive predictor of quiz ($p = 0.006$) and lab assignment ($p = 0.002$) achievements. The unstandardized coefficients for the total frequency of answering H5P questions were 1.404 and 0.842 for quiz and lab assignment achievements, respectively. Each one-unit increase in the total frequency of answering H5P questions is an increase in the total quiz score of 1.404 points. Similarly, each one-unit increase in the total frequency of answering H5P questions is an increase in the total lab assignment score of 0.842 points.

Table 4. Multiple regression analysis results for quiz, capstone, and lab assignment achievements

	Unstandardized Coefficient B	Unstandardized Coefficient Std. Error	Standardized B	t	p
Quiz success					
Total frequency of H5P questions	1.404	0.495	0.649	2.838	0.006
Total time spent on H5P questions	-2.849e-5	0.000	-0.035	-0.308	0.759
Total points awarded for H5P questions	-0.311	0.515	-0.137	-0.603	0.549
constant	65.380	4.355		15.013	< 0.001
Capstone success					
Total frequency of H5P questions	0.772	1.263	0.160	0.612	0.543
Total time spent on H5P questions	0.000	0.000	-0.073	-0.558	0.579
Total points awarded for H5P questions	0.380	1.315	0.075	0.289	0.774
constant	74.195	11.115		6.675	< 0.001
Lab assignment success					
Total frequency of H5P questions	0.842	0.258	0.722	3.265	0.002
Total time spent on H5P questions	-5.662e-5	0.000	-0.127	-1.154	0.253
Total points awarded for H5P questions	-0.201	0.269	-0.164	-0.747	0.458
constant	15.899	2.270		7.003	< 0.001

4. DISCUSSION

The result of this study is important to the literature for two reasons, a) deciding possible assessment types in an online course design and b) providing insight related to the parts of energy content that geoscience students monitor their learning through self-assessment resources. Integrated quizzes in the learning management system enabling engagement with the course content created a significant difference between those in the 1-credit degree completion course and the 3-credit course. This result made sense when we think that students in the 1-credit course engaged more with the content to complete their undergraduate degree. This coincides with the quizzes being the source of engagement with the learning throughout a course (Rocco, 2007; Taufiq-Hail et al., 2021).

Capstone projects enable students to learn actively in a contextual manner (Kearns, 2012). In this study, the capstone project allowed students to take responsibility for their learning and learning process. Quizzes enabling highly structured course design, capstone projects providing opportunities for contextual learning, and lab assignments were significant positive predictors of the overall success of the course. Integrating these different styles of assessments might allow students to engage with the course content more, particularly considering the asynchronous nature of the course.

Frequently completing H5P questions had a significant positive effect on quiz and lab assignment achievements. Completing the H5P questions more frequently in Module1 compared to other modules may have been due to students' curiosity about what H5P was and/or they thought they would need help and feedback at the beginning of the course. Answering H5P questions in the more specific and detailed content of Module3 and Module4 compared to Module 2 – What is Energy? indicates the supportive role of H5P questions. These findings support the role of ungraded self-assessment questions, which are just-in-time and delayed feedback along with learning support (Lewis et al., 2010). The reason for spending more time on H5P knowledge check questions in Module4: Global Warming – Physics compared to Module5: Global Warming – History might be the use of numerical tools and publicly available data for global warming in Module4 and less focus on numbers and data in Module5.

5. CONCLUSION AND RECOMMENDATIONS

Quizzes, lab assignments, and the capstone project were assessment tools that positively affected overall success. The total frequency of answering H5P questions was a significant positive predictor of quiz and lab assignment achievements. Course and learning designers might consider adding various types of assessment strategies to increase students' overall achievements in an online course. H5P questions were used more frequently in more specific and detailed concepts in this online geoscience course. Integrating self-assessment questions to more specific and detailed modules would be more supportive to students. Designers and instructors would only create and integrate self-assessment questions to more specific and detailed modules, in such way, they save time and reduce their workload. For future course design, discussions and lab assignments in each module can be considered in addition to the quizzes and the capstone project for student engagement and model simulations. This would be more aligned with the recommendations regarding online assessment during the COVID-19 pandemic, particularly related to debates amongst students and continuous assessment (García-Peñalvo et al., 2021). Also, the number of participants in this study was limited. For further research, other statistical tests can be conducted with more participants' data to test the indirect effect of H5P questions through other assessment types on overall success.

REFERENCES

Ally, M., 2008. Foundations of Educational Theory for Online Learning. In T. Anderson (Ed.), *The Theory and Practice of Online Learning* (2nd ed.), pp. 15–44. Athabasca University Press.

Freeman, S. et al., 2011. Increased Course Structure Improves Performance in Introductory Biology. *CBE—Life Sciences Education*, Vol. 10, No. 2, pp. 175–186. https://doi.org/10.1187/cbe.10-08-0105

García-Peñalvo, F. J. et al., 2021. Recommendations for Mandatory Online Assessment in Higher Education During the COVID-19 Pandemic. In D. Burgos, A. Tlili, & A. Tabacco (Eds.), *Radical Solutions for Education in a Crisis Context*, pp. 85-98. Springer. https://doi.org/10.1007/978-981-15-7869-4_6

Haak, D. C. et al., 2011. Increased Structure and Active Learning Reduce the Achievement Gap in Introductory Biology. *Science*, Vol. 332, No. 6034, pp. 1213–1216. https:///doi.org/10.1126/science.1204820

Hartwig, M. K. and Dunlosky, J., 2012. Study Strategies of College Students: Are Self-testing and Scheduling Related to Achievement? *Psychonomic Bulletin & Review*, Vol. 19, No. 1, pp. 126–134. https://doi.org/10.3758/s13423-011-0181-y

Kearns, L. R., 2012. Student Assessment in Online Learning: Challenges and Effective Practices. *MERLOT Journal of Online Learning and Teaching*, Vol. 8, No. 3, pp. 198–208. https://jolt.merlot.org/vol8no3/kearns_0912.htm

Lewis, D. et al., 2010. Knowledge Check Questions: Best Practices for Use of This Instructional Strategy. In J. Herrington & C. Montgomerie (Eds.), *ED-MEDIA 2010—World Conference on Educational Multimedia, Hypermedia & Telecommunications*. Association for the Advancement of Computing in Education, pp. 2783–2788. https://www.learntechlib.org/primary/p/35034/

Montenegro-Rueda, M. et al., 2021. Assessment in Higher Education during the COVID-19 Pandemic: A Systematic Review. *Sustainability*, Vol. 13, No. 19. https://doi.org/10.3390/su131910509

Motade, S. and Deshpande, A. 2022. Active Learning Techniques for Effective Online Teaching and Learning in Higher Education. *Journal of Engineering Education Transformation*, Vol. 35, No. 4, pp. 112-120.

Nguyen, D. M. et al., 2006. The Impact of Web-based Assessment and Practice on Students' Mathematics Learning Attitudes. *The Journal of Computers in Mathematics and Science Teaching*, Vol. 25, No. 3, pp. 251-279. https://www.learntechlib.org/primary/p/6277/

Raes, A. et al., 2020. Learning and Instruction in the Hybrid Virtual Classroom: An Investigation of Students' Engagement and the Effect of Quizzes. Computers & Education, Vol. 143. https://doi.org/10.1016/j.compedu.2019.103682

Rocco, S., 2007. Online Assessment and Evaluation. *New Directions for Adult and Continuing Education*, Vol. 113, pp. 75–86. https://doi.org/10.1002/ace.249

Salas-Morera, L. et al., 2012. Analysis of Online Quizzes as a Teaching and Assessment Tool. *Journal of Technology and Science Education*, Vol. 2, No. 1, pp. 39–45. http://hdl.handle.net/2099/12152

Şahin, M. et al., 2020. Sequential Analysis of Online Learning Behavior According to E-Learning Readiness. In P. Isaias, D. G. Sampson, D. Ifenthaler (Eds.), *Cognition and Exploratory Learning in the Digital Age (CELDA)*. International Association for Development of the Information Society, pp. 117-131.

Taufiq-Hail, G. A. et al., 2021. Academic Staff Perceptions on Students' Traditional Assessment Transformations toward Online Evaluation during COVID-19 Pandemic in Higher Education: A Preliminary Study from Two Diverse Cultures. *2021 Sustainable Leadership and Academic Excellence International Conference*. Manama, Bahrain. IEEE, pp. 16-25. https://doi.org/10.1109/SLAE54202.2021.9686904

Wilkie, S. et al., 2018. Considerations for Designing H5P Online Interactive Activities. In M. Campbell, J. Willems, C. Adachi, D. Blake, I. Doherty, S. Krishnan, S. Macfarlane, L. Ngo, M. O'Donnell, S. Palmer, L. Riddell, I. Story, H. Suri, & J. Tai (Eds.). *ASCILITE*, pp. 543–549.

Yan, Z. and Brown, G. T. L., 2017. A Cyclical Self-assessment Process: Towards a Model of How Students Engage in Self-assessment. *Assessment & Evaluation in Higher Education*, Vol. 42, No. 8, pp. 1247–1262. https://doi.org/10.1080/02602938.2016.1260091