UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

The Role of Formative Assessment in a Blended Learning Course

Klinkenberg, S.

DOI
10.1007/978-3-319-57744-9_6

Publication date
2017

Document Version
Final published version

Published in
Technology Enhanced Assessment

License
Article 25fa Dutch Copyright Act (https://www.openaccess.nl/en/policies/open-access-in-dutch-copyright-law-taverne-amendment)

Link to publication

Citation for published version (APA):
Klinkenberg, S. (2017). The Role of Formative Assessment in a Blended Learning Course. In D. Joosten-ten Brinke, & M. Laanpere (Eds.), *Technology Enhanced Assessment: 19th International Conference, TEA 2016, Tallinn, Estonia, October 5-6, 2016 : revised selected papers* (pp. 63-71). (Communications in Computer and Information Science ; Vol. 653). Springer. https://doi.org/10.1007/978-3-319-57744-9_6

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:12 Oct 2026

The Role of Formative Assessment in a Blended Learning Course

Sharon Klinkenberg(✉)

Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129B, 1001 NK Amsterdam, The Netherlands
S.Klinkenberg@uva.nl

Abstract. Assessing the effectiveness of a course design in higher education is an almost unfeasible task. The practical inability to conduct randomised controlled trials in a natural setting limits teachers who want to interpret their design choices to non causal evaluation. With at least a moderate amount of common sense these evaluations could be useful in gathering insight into what works and what does not work in education. In our blended learning course we wanted to assess the role of formative assessment while also taking lecture attendance into account. There are certainly many confounds for which we cannot control. We found no effect of lecture attendance but formative assessments did predict a substantial amount of course outcome. This was probably due to not including the formative performance in the final grading and our emphasis in the course that to err is to be student. The validity of the formative assessment paves the way for diagnostic use and remedial teaching.

Keywords: Formative assessment · Blended learning · Lecture attendance · Weblectures

1 Introduction

In this short paper we would like to explore the role of formative assessment in a blended learning course. Though formative assessment has a broad scope [1], we focus on online trail assessment. For the last three years a new course has been taught at the department of Psychology of the University of Amsterdam. The course Scientific and statistical reasoning combines a broad array of online and offline learning methods. We will describe the course design, scope, place in the curriculum and our attempt to assess the effectiveness of formative assessment within this course.

Blended learning can be interpreted and implemented in many ways. The interplay of many variables eventually determine a specific course outcome. Singling out one variable is therefore not easily done without a decent experimental design. However, very little time is available for lecturers in higher education to assess the methods they use. Although the generalizability of such an attempt is limited, it does provide some insight into best and worst practises. In this paper we therefore will present our results on data that is available to us while acknowledging the limited methodological robustness.

© Springer International Publishing AG 2017
D. Joosten-ten Brinke and M. Laanpere (Eds.): TEA 2016, CCIS 653, pp. 63–71, 2017.
DOI: 10.1007/978-3-319-57744-9_6

1.1 Course Design

The course Scientific and statistical reasoning is taught in the second bachelor year. Consisting of 15 ECTS (European Credit Transfer and Accumulation System) during the first semester for a body of about 400 students, it can be considered a large course according to our university standards. The course consists of four distinct cycles of four weeks each concluded with an exam in the fifth week. Figure 1 illustrates how one cycle of five weeks is designed. It also shows that students have the ability to attend a walk-in session in which specific individual issues can be addressed. Furthermore, the course combines several traditional and modern methods. There are several weekly lectures and mandatory study groups. Prior to each study group meeting an assignment needs to be uploaded to the campus electronic learning environment (ELO) and every week a digital formative assessment needs to be taken. The assessments consist of assignments in which instruction and assessment are integrated. For example, students are instructed on the subject of moderation analysis. They are presented with a case and data files which need to be used to answer the presented questions. Figure 2 illustrates such an assignment. On top of the exam students also need to write a significant argumentation essay (AE) on a specific subject.

All the different ingredients of the course are integrated in the final grading scheme by means of the following formula (1), where the final grade is determined by the mean on all four exams, the mean on the argumentation essays and deduction points.

$$\text{Grade} = .8 \times \overline{\text{Exams}} + .2 \times \overline{\text{AE}} - \text{Deduction} \tag{1}$$

Deduction points are introduced when students do not meet the set deadlines for the formative assessment, the preparatory assignments for the study groups and study group attendance. The deduction points are presented in a graduated fashion, as can be seen in Table 1. The table shows that deduction will only take effect when multiple deadlines are not met and is created to be a reasonable

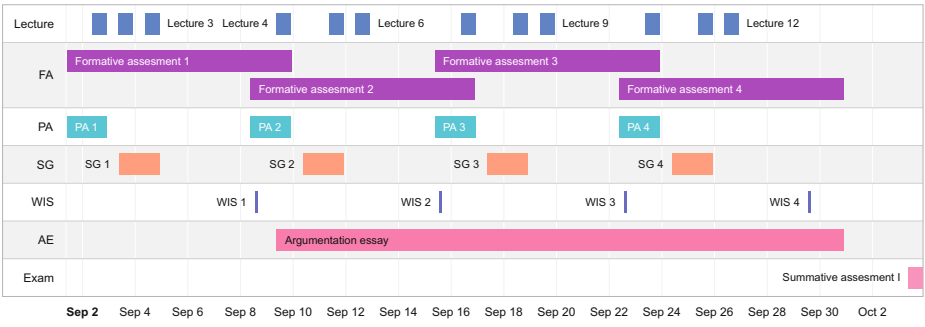

Fig. 1. Course design of cycle 1 of 4, for scientific and statistical reasoning. Showing lectures, formative assessments (FA), preparatory assignments (PA), study groups (SG), walk-in sessions (WIS), the argumentation essay (AE) and the mid term exam.

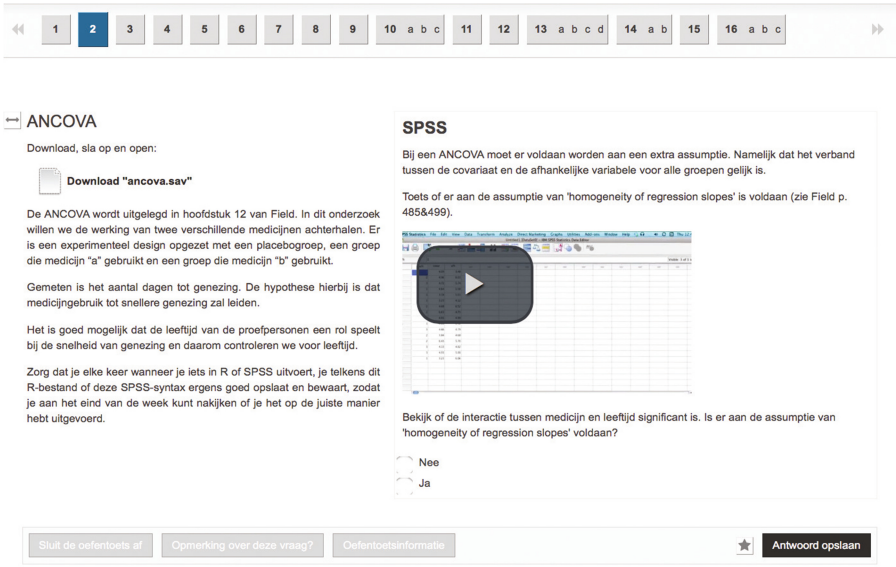

Fig. 2. Example of formative assessment question.

Table 1. Deduction points for deadlines not met. For the study groups (SG), prepatory assignments (PA) and Formative assessments (FA).

Not met	SG	PA	FA
1x	0	0	0
2x	0.25	0.1	0.1
3x	0.75	0.25	0.25
4x	1.5	0.5	0.5
5x	3	1	1
6x	Exclusion	Exclusion	Exclusion

penalty. It appeals to the phenomenon of loss aversion wherein Tversky and Kahneman [6] postulate that people tend to modify behaviour more for the avoidance of loss than for the gain of profit.

For both the formative assessment and the mandatory study groups with their prior assignment we only registered if the task was done. The quality of the work is not incorporated in the final grade. It is generally emphasized that students are allowed to make mistakes in their work but that their effort is taken into account. In recent years we encountered massive harvesting of our formative assessment items due to the fact that performance marginally determined the final result. Even our good students who did spend time on the assignments resorted to looking up the correct answer before submitting the final answer. The results on the formative assessment did not show any predictive validity with

the course exams and therefore could not be used as a diagnostic instrument for assessing individual problems in certain subdomains. We emphasize that making errors is part of the learning process and that with sufficient effort the course objectives can certainly be met. With this approach we aim to appeal to a growth mindset [3] that states that people who believe ability is not rigid have less problems obtaining new knowledge.

Apart from the above described course design, students also have access to the weblectures from our campus video lecture capture system. These weblectures are available directly after the lecture has been given.

2 Methods

In our analysis we aim to identify the role of formative assessment in the scope of the above described blended learning course. There are a lot of confounds and we can only present results on the limited variables we have access to.

The variables that determine the final course grade e.g. exams and argumentation essays are used as validation criteria. The deduction points are not taken into account as they indirectly determine the correlation between course outcome and predictors and therefore violate conditional independence. This final course grade is a non unidimensional indication of course performance. At least two kinds of skills can be theoretically extracted, namely the ability to perform and interpret statistical analysis and general scientific reasoning ability. For the statistics part SPSS is taught and R [5] is used for manual calculations. Furthermore, both the exam items as well as items used in the weekly assignment are tagged on subcategories. We use the scores on these subcategories to inform students on their performance.

2.1 Exam Results

There are four midterm exams which together determine 80% of the final grade. An exam consists of a statistics part and a critical thinking part and are administered digitally. An entire exam is constructed of multiple choice questions, fill in the blank questions and small essay questions. Furthermore, students are required to download data files and run analysis in SPSS and interpret the output from said analysis. Figure 3 shows an exam item as used. The results of the exams are subjected to general quality control analysis and corrected when necessary. Finally, items are tagged to categories and these sub scores are made available to students.

2.2 Mandatory Tasks

The study group attendance, preparatory assignments and formative assessments are all mandatory and result in course exclusion if deadlines are not met. They are scored as 0 or 1 and are incorporated as deduction points in the final grade as shown in Table 1. The inclusion of these binary results will not add any

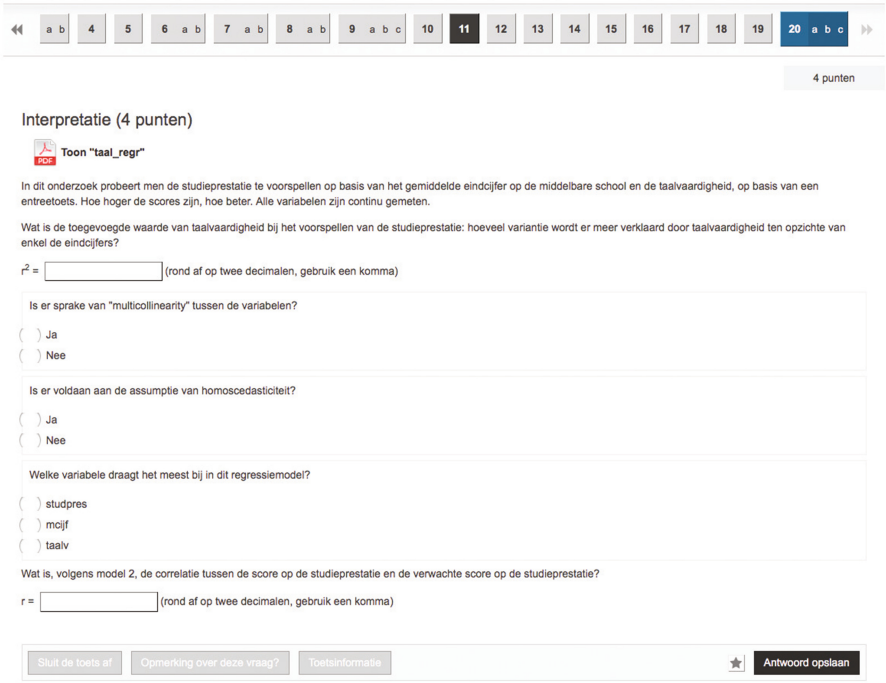

Fig. 3. Example of summative assessment item.

predictive value in course performance due to the fact that all students comply and only some are excluded from the course. We therefore will not add these binary variables to our model.

2.3 Formative Assessment

Apart from the binary compliance score the formative assessments also produce a score indicating performance about the whole assessment and also on all subcategories. These results will be taken into account. For the entire course there are 14 of these assessments and a general formative assessment indicator is created by extracting a factor using a principal component analysis (PCA). We used this method to arrive at an optimally fitting aggregated assessment facter with minimal bias.

2.4 Lecture Attendance

Though lecture attendance was not compulsory we did register course attendance in the first and final cycle of the course. This was done to get some indication of attendance drop during the course like we had experienced in the previous year. Students were asked to submit their student id at the beginning of the lecture.

We scored the attendance for every student as zero or one. Again we used the PCA method to arrive at a single factor for attendance.

2.5 Weblectures

Only the anonymous data on weblecture viewing behaviour was available to us. We therefore were not able to relate course performance to this variable. Our department feared that students whom did not attend campus lectures would procrastinate and would under achieve in this course. We analyse the anonymous results to get an insight into viewing behaviour.

In our analysis we will look at the formative assessment results to the lecture attendance, furthermore we will fine-grain the predictive validity of the formative assessment and finally we will eyeball the weblecture viewing behaviour.

3 Analysis

Our analysis is based on the $N = 427$ (70% female) students that attended the course. The average age for males and females was about 21 (sd males 2.14, sd females 2.7). Where the youngest student was 18 and the oldest 47. By far most students were 20 years of age.

We will first take a general look at the variables we can use to predict course performance which are lecture attendance and formative assessment results. There were 14 formative assessments during the course and we registered the attendance on 15 lectures, of which all lectures given in cycle 1 and all lectures in cycle 4. To reduce the amount of data we used a principal component analysis to extract one factor for course attendance and one for the formative assessment. Both factors were used in a regression analysis to predict the final grade corrected for the deduction points. This model explained 29% $F(2, 271) = 55.2, p < 0.001$ of the variance but looking at the beta coefficients revealed lecture attendance did not contribute to the model $b = -0.006$ $t(271) = -0.31, p > .05$. The coefficient for the formative assessment scores was $b = 0.22$ $t(271) = 10.49, p < .001$ (Eq. 2). For this regression analysis 150 observations were dropped due to missing values. This resulted from the PCA not being able to retrieve a factor score due to lack of variance or subjects not having a score on one of the three variables used.

$$\widehat{\mathrm{FG}} = 6.55 - .006\,\mathrm{LA} + .22\,\mathrm{FA} \tag{2}$$

It is fair to assume that lecture attendance in cycle one and four could not contribute to the total course performance because attendance was not measured in the second and third cycle. Therefore, we also analysed the predictive value on both predictors solely on the first and final exam. Here we used all course attendance results for the first cycle to predict the grade at the first midterm exam. Again attendance explained no variance $R^2 = .03$ while the four formative assessment results explained 24% $F(380) = 31.7, p < .001$. Cycle four showed that both attendance and formative assessment did not explain the grade for

Table 2. Regression R^2 for predicting final grades (FG) and mid term grades (MT) one to four with lecture attendance (LA) and/or formatieve assessment (FA) in the model.

Model	MT1	MT2	MT3	MT4	FG
LA + FA					0.28
LA	0.03			0.05	
FA	0.25	0.23	0.10	0.05	

the fourth midterm exam. Both models explained less than 5% of the variance. Though both were significant the effect is negligible.

For cycle two and three no lecture attendance data was available. We therefore only analyzed the results of the formative assessments. These respectively explained 23% $F(4, 368) = 27.74, p < .001$ and 10% $F(4, 374) = 10.89, p < .001$ of the variance. Table 2 summarizes the explained variance per model. Reliability of the four exams ranged from .53 to .68 on Cronbachs alpha.

3.1 Weblectures

It can be assumed that lecture attendance did not contribute to course outcome due to the fact that many students resorted to the weblectures. A combination of weblecture and attendance data would probably show that at least viewing the lectures would contribute to performance. Due to the privacy policy of the university we were not able to obtain the personalised results, therefore this combined insight could not be attained. As stated, the drop in lecture attendance had the department worried about course outcome due to procrastination. If this were the case we would expect the viewing frequencies to spike in the week leading to the exams. Figure 4 clearly shows the viewing frequency to be centred within the week of the lecture. The graph displays the frequency per day for the entire duration of the course for all given lectures. We only see a slight dent in the week leading up to the exam. We can therefore conclude that procrastination is no issue within this blended learning course.

4 Discussion

The results show that formative assessment plays a notable role in predicting course performance, though not all cycles contributed equally well. In particular the final cycle showed poor performance. We attribute this to the limited amount of formative assessments in this final cycle. Compared to the rest of the course, only two assessments were administered. Another component here is that the final exam also consisted of items containing topics on the previous cycles. The amount of lectures was also the lowest in this final cycle, though here the argument of exam content does not hold.

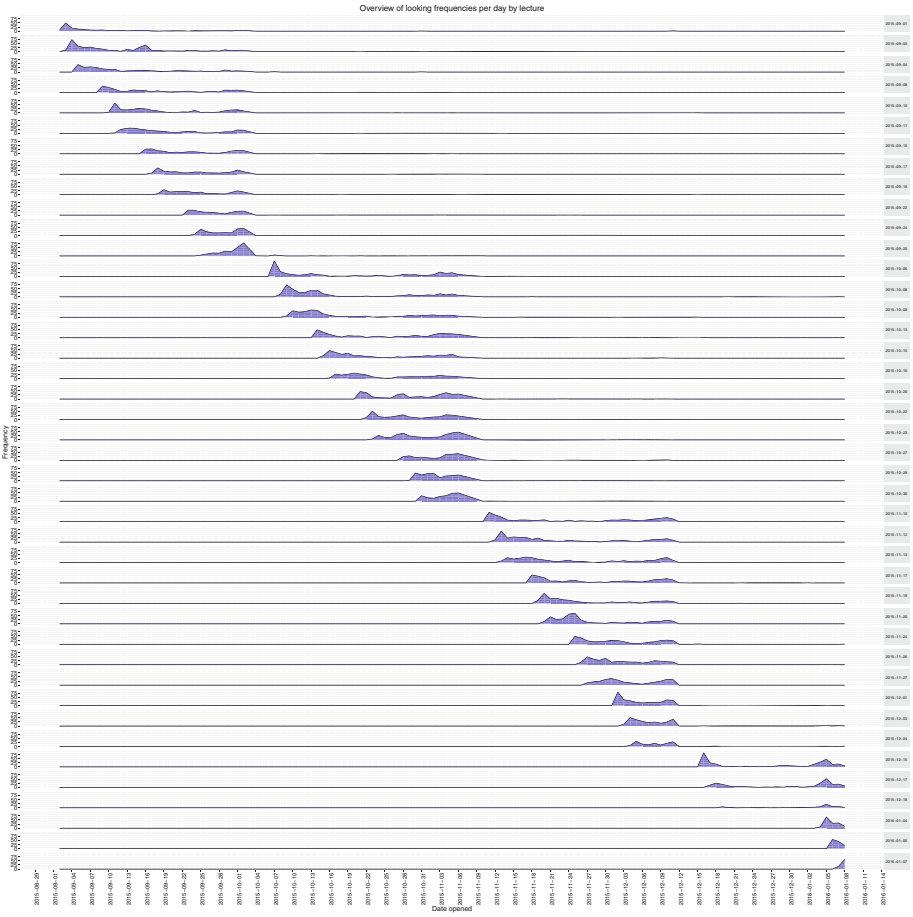

Fig. 4. Weblecture viewing frequencies for all lectures y-axis and course duration x-axis.

As stated earlier, the lack of any predictive value for lecture attendance could be because the weblecture data was not taken into account. It would be reasonable to at least assume that viewing the lecture content would influence course outcome.

That assumption does lead us to the question if attendance equals retention of content. Research on retention by Freeman et al. [4] at least shows that lectures without any form of active learning are not that effective. Our panel group discussions with students also indicate that not all students benefit equally well from lectures. Some find it hard to concentrate or to keep track. Others mention being distracted by other students. On the other hand students report the weblectures to create some differentiation. Our top students mention not attending the lectures and opting to watch the weblectures at double speed, while our

struggling students use the weblectures to revisit the content more often. It is therefore not unreasonable to expect weblecture views to be a better predictor of course outcome than lecture attendance. It should be strongly noted that this expectation only holds in the context of a blended learning setup in which pacing is highly controlled. Though research by Bos and Groeneveld [2] on the effects of video lectures show no difference compared to lecture attendance.

5 Conclusion

The results of this analysis show that formative assessment has a decent amount of predictive value when it comes to course outcome. Not incorporating the performance on the formative assessment in the calculation of the final grade, and emphasizing that errors are allowed seem to have worked to restore predictive validity. It can also be concluded that a sufficient amount of formative assessment needs to be administered and that the exams need to be properly aligned with the formative assessment. In the coming academic year we will start using the formative assessments as a diagnostic tool to facilitate remediation.

References

1. Black, P., Wiliam, D.: Developing the theory of formative assessment. Educ. Assess. Eval. Account. **21**(1), 5–31 (2009)
2. Bos, N., Groeneveld, C., Bruggen van, J Brand-Gruwel, S.: The use of recorded lectures in education and the impact on lecture attendance and exam performance. Br. J. Educ. Technol. (2015)
3. Dweck, C.S.: Self-theories: Their Role in Motivation Personality and Development. Psychology Press, Philadelphia (2000)
4. Freeman, S., Eddy, S.L., Mcdonough, E., Mcdonough, M., Smith, M.K., Okoroafor, N., Jordt, H., Wenderoth, M.: Active learning increases student performance in science, engineering, and mathematics. In: Proceedings of the National Academy of Sciences of the United States of America (2014)
5. R. Core Team: R: a language and environment for statistical computing (2016)
6. Tversky, A., Kahneman, D.: The framing of decisions and the psychology of choice. Science **211**, 453–458 (1981)