MEASURING EFFECTS OF VIDEO LECTURES ON IMPROVING STUDENT ENGAGEMENT AND OUTCOMES

V. Engbers
Eindhoven University of Technology
Eindhoven, the Netherlands

R. Gündlach
Eindhoven University of Technology
Eindhoven, the Netherlands
0009-0006-4204-4417

M. Regis
Eindhoven University of Technology
Eindhoven, the Netherlands
0000-0003-4306-8673

M. Vlasiou[1]
University of Twente / Eindhoven University of Technology
Enschede / Eindhoven, the Netherlands
0000-0002-0457-2925

Conference Key Areas: *Innovative Teaching and Learning Methods, Virtual and Remote education in a post Covid-19 world*

Keywords: *Video Integration, Lesson Activities, Improvement Grades, Post-Pandemic Education, Digital Education, Mathematics Master Course*

[1] Corresponding Author:
Maria Vlasiou, m.vlasiou@utwente.nl; m.vlasiou@tue.nl

ABSTRACT

In the current age, digital advancements have shaped the educational landscape by providing numerous possibilities for a fast and on-demand influx of information for students. This brings an additional difficulty for course designers in how to incorporate such technologies in teaching in an optimal way. Key examples are educational videos, which are especially relevant now due to the increase in accessibility of pre-made videos and recording technology since the pandemic. This puts post-pandemic teaching in the new but revolutionary position to easily complement in-person teaching with such videos.

In this study, we examine the effect of videos combined with in-person teaching in a mathematics master course in motivation and grades. This experiment is specifically insightful due to our course consisting of three different topics (A,B,C). In Year 1 (control group), the course was taught traditionally. In Year 2 (experimental group), we provided additional video lectures on (A), while keeping (B) and (C) as before. We compare assessment and survey results between and within years.

Videos did not increase the students' motivation for the topic (Fisher exact test p = 0.06182). The intervention also did not improve the midterm or final exam grades on (A) between years. Students who watched videos did not score significantly better on their assignments (2MWUt p = 0.275) nor on their exams (2MWUt p = 0.745) than students who did not watch the videos. However, a positive effect size was observed between years, while the intervention led to a negative effect size within the same year.

1 INTRODUCTION

Technology is fully integrated into everyday life, which has had a positive effect on students' internet and computer skills and their attitude towards digital educational resources from a young age (Kuhlemeier and Hemker 2007) and (Sharples et al. 2007). This presents an opportunity in transferring this experience in a classroom environment. Specifically for learning mathematics at university, using *online videos* has been shown to have advantages. The topic of video integration into in-person lectures is especially relevant in a post-pandemic teaching environment, where copious amounts of video material and recording hardware, used during the pandemic, are available. Teaching staff both acknowledges the importance of in-person teaching and recognises the benefits of videos and recordings (Robson, Gardner, Dommett 2022). It is therefore of great interest to investigate if the combination of in-person teaching and video material can elevate teaching in a post-pandemic time.

Benefits of videos include flexibility of scheduling and pace, and avoidance of long lectures. In contrast, the main perceived advantages of lectures are the ability to engage in group tasks, to ask questions, and to learn 'gradually' (Howard, Meehan, Parnell 2018). The same study also shows that students in clusters with high lecture attendance achieve, on average, higher marks in the module. Therefore, videos provide a useful resource, which should be used in this context only and in conjunction with lectures. Further studies focus on the relationship between learning and time spent on lectures and/or videos. Findings show that students use videos as either a complement to or substitute for the lecture, and time spent using either or both resources has a significant impact on learning (Meehan and McCallig 2019).

In this work, we measure the effect of video-integrated education in combination with in-person teaching. The goal of this work is to answer the following questions:

1. Does the addition of videos to in-person teaching improve the students' grades?
2. Does the addition of videos to in-person teaching improve the students' engagement with the course?

In order to give a statistically sound answer to these questions, we track the data of two years of the same master course in mathematics at the Eindhoven University of Technology in the Netherlands. In the first year, the course was given traditionally, while in the following year videos were included. The first year serves as a benchmarking iteration where typical course standards are maintained. This is followed by an experimental iteration in the following year where, in addition to the standard practice of the course, video lectures are *partly* provided, i.e. *only* for a specific part of the course. This creates an additional natural control group. It allows us to compare the effect of the videos by comparing grades of the topic where videos were provided for, while controlling for confounding factors (e.g. the assumption that a year may have had stronger students) by comparing grades between years on the topics without videos. We refer to Section 2.1 for a detailed overview of the experimental setup.

We use the answers to these questions to formulate concrete advice if additional videos have an inherent added value to in-person teaching for students. Based on this advice, one may include videos in their course if the expected benefits justify the investment of time and resources. Note that these conclusions are in the specific context of a specialized and rather challenging first year master course in mathematics.

2 METHODOLOGY

2.1 Experimental setup

The experiment takes place in an established master course in mathematics in the Netherlands. The course comprises three topics in stochastics: renewal processes (A), branching processes (B), and Brownian motion (C), which are taught sequentially in three modules in this order. The three topics are sufficiently independent; i.e. for any module there are no required prerequisites from a previous module. Throughout the course, basic probability topics (D) appear as needed. The design involves a two-hour, on-campus lecture, followed by a two-hour guided self-study, twice per week for eight weeks. Examination weeks are scheduled afterwards. The material offered to students is lecture notes, one single book covering all topics, instruction sets, sketches of solutions to the instruction exercises, and practice exam sets for the final examination.

To assess students, at the end of each module a midterm examination is given in the form of take-home problems that can be solved in pairs or alone. At the end of the course, students also take a final exam where all topics are tested in a three-hour individual examination. The midterms and the final exam are graded on a 0–10 scale. The three midterms count each for 10% of the final grade and the final exam counts for 70% of the final grade. A final grade of at least 5.50 is needed to pass the course.

The course was taught with the same setup, teaching staff, and material (with the exception being the inclusion of the videos required to perform this experiment) between two years. In Year 1, the course was given traditionally, i.e. with the setup described above. In Year 2, videos were recorded by the lecturer on (A) and offered to all students at the beginning of the course. Students could engage with the videos throughout the whole course and examination period, i.e. for a total of 10 weeks.

The videos used in Year 2 consisted of six mini-lectures: three informative presentations on theoretical topics and three on applications. For both parts, one video was in the scope of (A) and handled in class and two were new material. The videos were put online and it was indicated if a video was on theory or on an application. The differentiation between theory and applications as well as known and new subjects could have allowed for a determination of whether students are seeking help when engaging with a video or are intrinsically interested to learn more about (A). However, this differentiation, together with the fact that the material was optional, reduced the relevant sample size per video, thus not allowing for meaningful statistical analysis.

By providing videos only for one out of three topics, a meaningful comparison can be made between the performance of the students in two consecutive years on the topics with and without videos. Students were offered the new video material but were not obliged to follow it. All students were asked to record whether they engaged with the videos or not. This created a natural control group also within Year 2, both within (A) and between topics.

While Year 2 took place during the Covid-19 pandemic, this specific course was chosen by the Programme Director for on-campus education allowing for a fair comparison between the two years. In addition to keeping all setup, material, and staff identical, care was taken to avoid confounding factors by design. For example, each midterm per topic and each topic in the final exam was assessed by the same person between both years. The number of registered students was 73 and 74 in the two years.

2.2 Data and reliability

We use anonymised data of students of Years 1 and 2 contained in three datasets. Dataset 1 consists of midterm grades of (A,B,C) and exam grades per topic (A,B,C,D) and year. The grades on (A) between years may attest for the effect of videos, while the grades on (B,C,D) give an indication of difference between the two years as no changes in education were made for these topics, thereby assessing the effect of possible confounding factors due to the different cohorts.

To perform statistical tests based on Dataset 1, we ensure that student groups from Years 1 and 2 are independent. Therefore, we remove students that were present in both years to avoid dependencies. For a fair comparison, we also remove students in Year 1 if they had taken the course the previous year, as they had prior knowledge. However, this introduces a bias toward higher grades in Year 1 and therefore students that scored lower than a 5 on their exams in Year 2 are removed as well. Therefore, the data is limited to the groups of students from either year that took the course for the first time *and* passed, making these groups independent. The resulting sample sizes in each year are again similar (45 and 42 students).

As is usual with test or survey data, we test the reliability of the midterm, exam, and survey questions in order to identify and remove questions that are not discriminatory. In Dataset 1, we measure the overall reliability of the midterm or exam with the Cronbach alpha coefficient (Cronbach 1951). We then look at the average score of each question and its correlation with the other questions in the exam. If the score is below 0.15 or above 0.85 of the total score on the question or the correlation with the other questions is below 0.15, the question is removed from the analysis. Data per question was not available for the midterm on (B) and (C) in Year 1 and thus all questions of these midterms are included in the analysis.

Dataset 2 consists of survey results held amongst students (on a voluntary basis) in Year 2. The first part of the survey consisted of basic questions: which of the six video lectures (if any) did they watch; which was their favourite; what was the number of hours they spent on the course per module. The latter information is used to investigate if students that watched at least one video are statistically different than those who did not watch the videos (e.g., harder or less hard working). This is a key measure to account for the important confounding factor of *selection bias* of the groups, since watching the videos was voluntary. Surveys were performed at the end of each module rather than only at the end of the course, thus allowing students to have a fresh (and hopefully accurate) estimation of the effort they expended. The second part of the survey asked students two main questions on what the effect of the videos was for them:

1. If it made (A) more interesting for them;
2. If it made the course as a whole more interesting for them.

On each question the students could respond with “yes”, “somewhat” and “no”. This information is used to test the motivation for (A) and for the course as a whole of students between those that watched at least one video and those who did not.

Student Evaluation of Teaching (SET) reports for the two years were also available. They were used to test whether students in Year 2 spent more time on the course.

2.3 Statistical methods

To answer the first main question on whether the video lectures had an effect on grades, we examine the following hypotheses:

H_0: The probability of a uniformly sampled grade from Year 1 being larger / smaller than a uniformly sampled grade from Year 2 is ½.
H_1: The probability of a uniformly sampled grade from Year 1 being larger / smaller than a uniformly sampled grade from Year 2 is not ½.

For this, we use the grades in Years 1 and 2 (Dataset 1) and employ the two-sided Mann-Witney U test (2MWU) (Mann and Whitney 1947). This non-parametric test investigates the locations of two independent samples by using ranks and is suitable for small-sample ordinal data, as is the case when examining grades. We also report the median and interquartile range (IQR) of the grades of the two years separately.

To answer the second main question on whether the video lectures had an effect on the engagement of students with the course, after testing for differences in engagement between years, we examine the following hypotheses:

H_0: The results between groups in Year 2 are from the same distribution.
H_1: The results between groups in Year 2 are not from the same distribution.

For the comparison of survey results within Year 2 (Dataset 2), we use the Fisher's exact test (Sprent 2011). This test is suitable for categorical survey data, i.e. where students can answer either "yes", "somewhat", or "no", from independent groups of students that did not watch the videos and the group of students that watched at least one video. The statistical tests are performed on a significance level of $\alpha = 0.05$ that is adjusted to correct for multiple testing where needed: when the final exam is considered both as a whole, and as a set of subparts comprising the topics (Bonferroni correction, $\alpha = 0.025$). We do not correct for multiple testing when considering midterms and the final on the same topic as we believe they can be seen as independent due to the different modalities (take-home/in-class, group/individual, with/without resources).

In all relevant cases, in addition to statistical significance, we also report the effect sizes. Unless the impact of the intervention is huge, a study of this size is unlikely to get a statistically significant result. Thus, reporting the effect adds nuance to the results. The effect size is a standardized, scale-free measure of the relative size of the effect of an intervention. It is particularly useful for quantifying effects measured on arbitrary scales. In education, if it could be shown that making a small and inexpensive change would raise academic achievement by an effect size of even as little as 0.1, then this could be a very significant improvement, particularly if the improvement applied uniformly to all students, and even more so if the effect were cumulative over time; see also (Coe 2002) for a discussion.

3 RESULTS

3.1 Between years

The analysis of the SET results between years revealed that all aforementioned groups of students (i.e. the complete cohorts or only the independent sets) spent on average the same amount of time on the course and on the same module between years.

In Table 1, we compare the grades of Year 1 to the grades of Year 2 in Dataset 1. For the midterms, we observe a significant difference in grades for only midterm (C), while the other midterm grades did not differ significantly between years. This may indicate that Year 2 had stronger students. This is, however, in strong contrast with final exam grades on (C) and (D), where students scored significantly lower in Year 2. It is notable that the effect of the intervention is positive for all midterms and the intervention topic (A) in the final exam, but negative for all other topics in the final exam.

Table 1: Summary statistics of the grades per topic for the midterms (A,B,C) and final exams (A,B,C,D) in Years 1 and 2. Non-discriminatory questions were removed when possible. The p-value corresponds to the 2MWU test between Years 1 and 2. Boldfaced values indicate that the null hypothesis was rejected under a significance level $\alpha = 0.05$ (0.025 where a correction for multiple testing was needed). The Cliff's Delta effect size and 95% confidence interval are given.*

Between years	Year 1 (control)			Year 2 (intervention)			p-value	Effect size	95% CI of the effect size
	n	Me-dian	IQR	n	Me-dian	IQR			
Midterm (A)	30	8.15	[7.18,8.88]	27	8.45	[7.70,9.50]	0.149	0.217	[-0.087,0.483]
Midterm (B)	30	9.00	[8.00,9.75]	27	9.10	[8.80,9.40]	0.445	0.118	[-0.199,0.412]
Midterm (C)	30	7.50	[7.00,8.50]	27	8.70	[8.06,9.71]	**<0.001**	**0.546**	[0.260,0.744]
Final (A)	45	4.00	[1.00,7.00]	42	5.38	[4.75,6.25]	0.033*	0.265	[0.003,0.493]
Final (B)	45	7.60	[6.00,9.60]	42	7.5	[6.56,8.30]	0.520*	-0.080	[-0.318,0.167]
Final (C)	45	8.67	[6.89,9.78]	42	7.29	[6.82,8.61]	**0.007***	**-0.338**	[-0.522,-0.123]
Final (D)	45	5.00	[3.33,7.33]	42	3.33	[2.08,4.33]	**0.011***	**-0.318**	[-0.524,-0.076]
Final total	45	7.02	[6.21,7.84]	42	6.61	[5.45,7.19]	0.082*	-0.217	[-0.430,0.018]

3.2 Within Year 2

Next, we compare students within Year 2, separated in two groups: students who watched at least one video and those who did not watch any videos. We examine the results on the two main questions of the survey. The results are presented in Tables 2 and 3. To test whether the number of students reporting an increased motivation for (A), we perform a Fisher exact test. There is not enough evidence to reject the null hypothesis of equal distribution (p-value = 0.06182). To analyse whether videos increased motivation of the course as a whole, we perform the same test on the frequencies reported in Table 3. The outcomes between the two groups differ (p-value = **0.0122**). Thus, we conclude that students feel that videos increased their motivation.

Table 2: Survey results on the question: "Do you feel that the videos made renewal processes more interesting for you? Or in lack of interest, did you appreciate the topic more?"

Increase motivation for (A)	Yes	Somewhat	No	Total
Watched a video	5	9	5	19
Did not watch a video	0	4	7	11
Total	5	13	12	30

Table 3: Survey result on the question: "Do you feel that the videos made the course any more interesting? Or do you feel that the videos were of any added value to the course?"

Increase motivation course	Yes	Somewhat	No	Total
Watched a video	9	6	4	19
Did not watch a video	0	6	6	12
Total	9	12	10	31

We next consider if students that watched the videos are significantly more engaged with the course. This is tested based on the number of hours students report to have worked on the course per topic. It is important to note that these numbers are self-reported by students. For all samples (i.e. per topic and group) the Shapiro-Wilk test does not reject the hypothesis that the hours reported follow a normal distribution (p-values ranging from 0.092 to 0.925). An F-test does not reject the hypothesis that the samples have the same variance. We thus assume normality of the data and employ a two-sided *t*-test with equal variances to compare the effort in hours between groups. The results are presented in Table 4.

Table 4: Summary statistics of the number of hours students spent studying in total per topic. Students are split into groups that did not watch the videos and that indicated to have watched at least one video. The p-value reported is based on a two-sided t*-test with equal variances. The Cohen's d effect size with Hedges g correction is given.*

Within Year 2	Did not watch any videos (control)			Watched a video (intervention)			p-value	Effect size	95% CI of the effect size
	n	Mean	S.d.	n	Mean	S.d.			
Topic (A)	27	30.15	12.04	19	37.42	11.90	**0.049**	0.596	[-0.009,1.202]
Topic (B)	27	24.89	11.42	19	27.58	11.49	0.437	0.231	[-0.364,0.826]
Topic (C)	27	25.48	11.47	19	31.26	13.54	0.125	0.460	[-0.141,1.061]
Total	27	80.52	32.07	19	96.26	33.73	0.116	0.472	[-0.129,1.074]

Students who watched a video spent more time on (A). The difference in the mean is roughly equal to the time required to watch all videos, which is a possible explanation. All other differences in effort were not statistically significant.

We also consider if the group that watched the videos scored significantly better on their midterm and final exams. As students were allowed to make their midterms in pairs, we test if the pairs in which *at least one student* watched a video performed better on their midterms than pairs in which *none of the students* watched a video. Additionally, we present the exam grades of students in Year 2 for students that did not watch any of the videos and for students that watched at least one video.

Table 5: Summary statistics of results of student pairs per topic in Year 2. All pairs are clustered into two groups: one in which neither student watched a video and one in which at least one student watched at least one video. We report the p-value of the 2MWU test between groups. Boldfaced values indicate that the null hypothesis is rejected under a significance level $\alpha = 0.05$. (0.025 where a correction for multiple testing is needed). The Cliff's Delta non-parametric effect size and its 95% confidence interval are also given.*

Within Year 2	None watched a video (control)			At least one student watched a video (intervention)			p-value	Effect size	95% CI of the effect size
	n	Median	IQR	n	Median	IQR			
Midterm (A)	11	8.60	[7.78,9.65]	16	7.70	[6.85,8.95]	0.275	-0.325	[-0.683,0.159]
Midterm (B)	11	9.00	[8.73,9.38]	16	9.10	[7.95,9.25]	0.107	-0.149	[-0.524,0.273]
Midterm (C)	11	9.35	[8.39,9.79]	16	8.35	[7.75,8.85]	**0.048**	-0.383	[-0.723,0.106]
Final (A)	22	5.38	[3.91,6.19]	16	5.38	[3.97,6.06]	0.745*	-0.065	[-0.421,0.307]
Final (B)	22	7.40	[5.85,8.23]	16	7.00	[5.88,7.85]	0.679*	-0.082	[-0.430,0.286]
Final (C)	22	7.41	[4.64,8.61]	16	7.09	[6.07,8.39]	0.801*	-0.051	[-0.403,0.313]
Final (D)	22	2.10	[1.00,2.55]	16	1.90	[1.00,2.70]	0.413*	0.045	[-0.331,0.410]
Final total	22	6.12	[4.97,6.58]	16	6.04	[4.71,6.71]	0.859*	-0.028	[-0.389,0.340]

We observe only a significant change in grades on midterm (C), where the groups that watched the videos scored significantly lower. Therefore, there is no significant change in grades of midterms on (A) or on the exam on (A). Note that in all cases except for topic (D), the effect size of the videos on the grades is negative.

4 CONCLUSION

There no statistically significant improvement in grades between Years 1 and 2. On topic (A) of the final exam, grades seem to be improved, but statistical significance is not reached after Bonferroni correction (p-value=0.033). On the other hand, the effect size is positive (0.265, 95%CI=[0.003, 0.493]). Also on midterm (A), the effect size is positive (0.217, 95%CI=[-0.087, 0.483]), although the result is not statistically significant (p-value=0.149).

As evidenced by the lack of significant unilateral changes in grades for (B,C) between years, we find that one cohort of students was not significantly overall performing better or worse than the other. Only on the midterm (C), we observed an increase in grades (Y1 7.50 [7.00, 8.50], Y2 8.70 [8.06, 9.71], p-value<0.001), but the results on the final (B,C,D) have negative effect sizes and reach statistical significance for (C,D).

Based on Table 5, we also do not observe a significant difference in grades between students that watched at least one video and students that did not watch any videos within Year 2. This indicates that selection bias did not play a significant effect in the analysis. Also, by the removal of non-discriminatory questions, the most important confounding factors in this analysis have been accounted for. Therefore, we find only a mild marginal effect of videos improving students' grades.

We find some evidence of improvement in motivation in the group of students who watched the videos. Concretely, there is no significant improvement in motivation for (A) based on Table 2 but there is a significant improvement in motivation for the course as a whole based on Table 3.

Moreover, Table 4 shows that students who watched the videos did not engage more in the course in terms of hours spent per topic. As the group that watched at least one video did not significantly work harder for the course than those that did not, according to Table 4, we find that the former group was not engaged with the course more than the latter in terms of hours spent on the course. This outcome shows that the group that chose to watch the videos are not inherently working harder, if measured purely by the amount of time they spend on the course. Therefore, the effect of videos on motivation is marginal, which concludes our second research question. On the other hand, it could be that students already inclined to the subject (and that need to spend less time to grasp the compulsory material), also engage with the videos. However, this is not visible in the grades.

Combining these results implies that the effect of including video material in combination with in-person lectures is modest. Our sample size was limited, especially when considering the complex framework that is analysed and the small effect sizes at hand. This may have prevented us from finding statistically significant results. For similar courses, it is therefore advisable to critically assess if the time and effort for the creation and integration of video is justifiable when the benefits are expected to be limited. On the other hand, the analysis shows no significant negative effect on students' grades or motivation. If videos are readily available and easily to implement, we find no evidence against including them as optional material; however, expectations should be managed accordingly.

5 DISCUSSION

When interpreting the results, it is important to keep in mind that the outcomes are based on two years of a specific master course in mathematics in the Netherlands. Courses of a different level or other university subjects may be better or worse suited for video integration. Additionally, students of different age, level, or field of study may respond differently to videos. The videos were made by the lecturer, with the input of students who took the course in the past years, in order to calibrate the potential interest students may have in the material. Professionally directed videos, or of a different format (in duration or media used), could have a different effect on the engagement of students. Finally, while the course is given in English and open to students from other countries, a large part of the students is Dutch. Different cultures may respond differently to videos.

While we carefully accounted for different confounding factors, it is possible some inevitable and hard to control effects were present on the background. An example is the pandemic that was making its uprise during Year 1, whereas the first pandemic wave was on its decrease during Year 2. Although both years provided on-campus teaching, the pandemic was in very different stages between years. This may have influenced the experiment. For example, students may have been more experienced with video lectures in Year 2 or may have had intrinsic variations in focus and motivation for coursework after the extended measures the pandemic required. As this course was selected for on-campus education, this may have created a welcome respite from isolation, which could be a confounding factor for engagement.

Another potential confounding factor relates to the content of the midterms and exams. They must be distinct between years. We accounted for discrepancies by removing non-discriminatory questions and keeping the graders the same between years. The assessments were designed by the same person. These measures are however not a perfect technique as some differences in level may still be present. Similarly, while the teaching staff did not change between midterms, it is possible that the quality of lectures or grading between years differed. We believe that these effects are mild as Table 1 does not show a convincing consistent difference in results between years.

Future years can expand upon the analysis here. First, the selection bias can be eliminated by making the videos and survey mandatory. However, it is not straightforward how to properly implement this in practice, i.e., how to reward or penalise students who did or did not watch the videos respectively. If future studies increase the sample size of students, it can additionally be tested which type of videos (i.e., on theory or applications) students prefer and if one version typically results in higher grades or motivation. This would add an additional research question to what *kind* of videos are optimal to use.

This research is not limited to the kind of videos only, but may also examine in a similar manner the effect of other properties such as video length, quality, and presenter. If future studies allow us to follow students for a longer time, we can test the long-term effects of videos on students. This adds an additional research question if videos improve the *learning retention* of students in practice. While these questions are of great interest, they are left for future investigation.

Overall, our research shows no statistical significant results, but modest effects of this intervention. In closing, it is important to highlight and emphasize the critical significance of not merely assuming the effectiveness of technological interventions in education. Instead, it is crucial to encourage rigorous educational research that enables more thoughtful and informed assessments of similar integrations.

6 ACKNOWLEDGEMENTS

This material is based upon work supported by the Innovation Fund of the 4TU Centre for Engineering Education in the Netherlands under reference IF2020-Vlasiou.

REFERENCES

Coe, Robert. 2002. "It's the Effect Size, Stupid: What effect size is and why it is important." Paper presented at the *British Educational Research Association Annual Conference 2002, Exeter, 11–14 September 2002.*

Cronbach, Lee J. 1951. "Coefficient alpha and the internal structure of tests." *Psychometrika* 16, (September): 297–334. https://doi.org/10.1007/BF02310555

Howard, Emma, Maria Meehan and Andrew Parnell. 2018. "Live Lectures or Online Videos: Students' Resource Choices in a First-Year University Mathematics Module." *International Journal of Mathematical Education in Science and Technology* 49, no. 4 (October): 530–553. https://doi.org/10.1080/0020739X.2017.1387943

Kuhlemeier, Hans, and Bas Hemker. 2007. "The Impact of Computer Use at Home on Students' Internet Skills." *Computers & Education* 49, no. 2 (September): 460–480. https://doi.org/10.1016/j.compedu.2005.10.004

Mann, Henry B. and Donald R. Whitney. 1947. "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other." *The Annals of Mathematical Statistics* 18, no. 1 (March): 50–60. https://www.jstor.org/stable/2236101

Meehan, Maria and John McCallig. 2019. "Effects on Learning of Time Spent by University Students Attending Lectures and/or Watching Online Videos." *Journal of Computer Assisted Learning* 35, no. 2 (November): 283–293. https://doi.org/10.1111/jcal.12329

Robson, Louise, Benjamin Garder, and Eleanor Dommett. 2022. "The Post-Pandemic Lecture: Views from Academic Staff across the UK." *Education Sciences* 12, no. 2 (February): 1–16. https://doi.org/10.3390/educsci12020123

Sharples, Mike, Inmaculada Arnedillo Sánchez, Marcelo Milrad and Giasemi Vavoula. 2007. "Mobile Learning." In *Technology-Enhanced Learning: Principles and Products,* edited by Nicolas Balacheff, Sten Ludvigsen, Ton de Jong, Ard Lazonder, Sally Barnes, 233–249. Springer Dordrecht. https://doi.org/10.1007/978-1-4020-9827-7_14

Sprent, Peter. 2011. "Fisher Exact Test". In *International Encyclopedia of Statistical Science*, edited by Miodrag Lovric, 524–525. Heidelberg: Springer Berlin. https://doi.org/10.1007/978-3-642-04898-2_253