MEDIA-ASSISTED UNDERGRADUATE TEACHING – PANEL EVIDENCE ON LEARNING OUTCOMES

Dr. Stephan H. Schmucker
Lecturer in Business Administration, University of Hamburg
School of Economics, Business and Social Sciences, Von-Melle-Park 9
D-20146 Hamburg
stephan.schmucker@wiso.uni-hamburg.de

Abstract: Electronic learning media experience ever wider use in tertiary education as a supplement to traditional lectures, and they come in an increasing number of formats. However, we lack conclusive evidence as to which format produces the best learning outcomes, especially if we consider the vast range of possible applications. This study presents a test of three different formats - audio only, voice-over slides, and slides plus a video of the lecturer - employed in a 'inverted classroom' undergraduate teaching context. The results suggest that the learning effects of the video format may not warrant the additional effort required for its production, leaving voice-over slides as the preferred medium of instruction. Perhaps more importantly though, the paper provides a detailed account of the experimental set-up - which is one that many lecturers will easily be able to replicate in order to ascertain which media format is most suitable in the teaching context at hand.

1 INTRODUCTION

In undergraduate teaching and elsewhere, the classic lecture is increasingly being supplemented with, or even replaced by, various other types of learning media, including video lectures. Such media are typically available on the internet, often via public platforms such as YouTube. This availability makes the learning media more widely accessible, hence we see calls for their increased use, for example in the political environment of German universities (Hamburger Bürgerschaft 2015). Also, the fact that such media can be consumed repeatedly makes them more learner-friendly than lectures, amongst a range of other benefits (Karnad 2013).

One teaching approach that strictly necessitates the use of non-traditional learning media is the so-called inverted (or flipped) classroom, a blended learning technique. The classroom is flipped in the sense that the instructional content is delivered via home learning, often aided by video lectures or other types of educational media, while the lecture time is instead used for a more in-depth exploration of the subject, to answer questions about the subject matter, for homework-style exercises and increased interaction with the lecturer as well as among the students.

Non-lecture learning media enjoy ever wider use (Fischer and Spannagel 2012) and come in increasingly diverse forms, ranging for example from simple audio files to voice-over slides, Khan-style videos, lecture capture, or elaborate, animated picture-in-picture videos. What we lack though, according to Chen and Wu (2015) and many others, is sufficient evidence as to which types of learning media are most suitable in a given educational setting (subject discipline, student body, etc.), where suitability is typically defined in terms of student acceptance and / or learning outcomes. Chen and Wu provide a brief survey of several studies of the performance effects of different types of learning media without arriving at a clear preference. Fey (2002)

explicitly investigates whether audio-visual learning media can achieve better learning outcomes than audio-only media but fails to find any significant effects. Interestingly though, her subsequent survey showed that the students *considered* audio-visual input to be more effective.

A lecturer who is planning to supplement a course with additional learning media, perhaps in order to 'flip' the classroom, will thus find little guidance as to the optimal type of learning media in the literature. Furthermore, existing studies will be of no value to her in conducting her own investigation to find out which format is best suited to her specific circumstances because – to the best of our knowledge – any insights generated so far have been derived from extensive experiments involving, for example, questionnaires, laboratory tests, non-standard learning or testing equipment, including even brainwave detection (Chen and Wu, 2015).

This study aims to help the hypothetical lecturer in two ways. First, we provide additional evidence on the differential performance effects and the students' evaluation of three different learning media: narration only; voice-over slides; slides plus video of the lecturer. The study was conducted during normal lecture hours in a second-year undergraduate course on Introductory Human Resource Management at the University of Hamburg. Based on a panel data set with a fixed effects approach, the results confirm Fey's (2002) warning that audio-visual media may not be worth the effort.

Second, if the lecturer is unwilling to directly adopt for her specific situation what little advice the literature provides, this paper presents an experimental design that is applicable to virtually any standard lecture setting with only modest additional effort required. In other words, by following the procedure we document, any lecturer can find out over the course of the first two or three lectures of a term which one from a set of alternative media formats works best in her setting and then concentrate on delivering that format henceforth.

The only technical prerequisite for conducting such an experiment is the availability of a set of classroom response devices or 'clickers' (see, e.g., Kundisch et al. 2013). Classroom response systems are experiencing ever wider use and have been associated with a number of benefits in teaching (Kay/LeSage 2009, Caldwell 2007, Schmucker 2015, Schmucker/Häseler 2015, Simpson/Oliver 2007). Most relevantly for our purposes, clickers allow a lecturer to conduct single or multiple choice tests. The questions and corresponding answer choices are shown to the students who then use the clickers to select and to transmit their choices. All responses are recorded electronically and later analysed. Clickers thus allow us to conduct performance tests much more conveniently than any paper-based or computer-based method.

The paper proceeds as follows: Section 2 exposes the research design. Next, the data are described, both in terms of descriptive statistics and in terms of a number of thoughts on data quality. Section 4 presents the results of the performance tests and of a subsequent survey of student opinions on the three media formats. Section 5 concludes with a summary and discussion of the results.

2 RESEARCH DESIGN

The greatest challenge in determining the influence of consuming different media types on learning outcomes lies in excluding, or holding constant, any potentially contributing factors that cannot be taken into account as explanatory variables in a regression analysis: the students'

general aptitude, their prior knowledge of the subject area, their individual skill at completing single choice tests under time pressure, their very participation in the tests (selection effects), etc. Our approach to this problem consists in creating and analysing panel data. The students were asked to participate in three tests (held incidentally during the Monday lectures in three consecutive weeks), in preparation for which they were furthermore asked to consume one of three types of audio-visual learning media on the subject of Introductory Human Resource Management:

- **Type A** is a series of three videos in which the screen is split between a video recording of the lecturer delivering a mini-lecture of roughly 8 to 10 minutes in a studio setting, and the lecture slides, which build up in step with the oral presentation.
- **Type B** videos are identical in content to Type A, but they lack the video image of the lecturer, i.e. they show only the slides accompanied by the audio track of the lecturer's narration, which is identical to the recording shown in Type A.
- **Type C** has no visual content but comprises only the audio track.

The video / audio files are available, at the time of writing, at https://lecture2go.uni-hamburg.de/veranstaltungen/-/v/17553 and associated websites.

In the lecture preceding the first test day, the setup and purpose of the experiment was explained to the students, and they were asked to consume at any time before the first test day a specific type of learning media. Each student was assigned to one of the three media types in a quasi-random fashion – via the last digits of their matriculation number. That way, we were able to avoid the students self-selecting their preferred media types, which would have invalidated the results. Immediately after the introductory lecture, the videos were uploaded to the University of Hamburg's 'Lecture2Go' platform at a site known to the students. At the end of each of the first two test days, the students were again assigned to a media type in such a way that those students who participated in all three test days would have consumed all three media types.

We thus obtain a panel data set with a time (test days) and a cross-section (students) dimension. This permits us to employ a fixed effects approach in estimating the performance outcome of media consumption: Every student who has participated in at least two test days and consumed at least two different media types is assigned an individual dummy variable. That way, all student-specific factors of influence that cannot be controlled for in the model are kept constant and we are able to examine the differential effect of the media types for each student. Furthermore, a set of dummy variables was also created for the test days to account for the possibility that the three tests were of unequal difficulty.

The actual test days were conducted in the following fashion: Clicker devices were distributed to all students, who were first of all asked to transmit the last five (out of seven) digits of their matriculation number so that their results can be matched with those of the other test days while maintaining a degree of anonymity. The students were then asked to indicate with their clickers:

- which media type(s) they had actually consumed in the course of the preceding week – choosing from media types A, B and C, plus all possible combinations thereof, plus "none",
- how many days since they had last consumed the media – selecting an integer number from 0 (consumption on the test day) to 7,

- and how many times they had consumed the media – indicating from one to four or "more" times.

Finally, the students were to answer five single-choice questions on each test day whose contents derived from the video tutorials. Three to five answer choices were presented for each question. The entire test consumed about 30 minutes of lecture time each Monday.

3 DATA

3.1 DESCRIPTIVE STATISTICS

A total of 174 tests were completed – to various degrees – by the students over the course of the three test days (the 'full sample'). Using the available information on matriculation numbers, we were able to identify 40 students who participated in two or three test days, i.e. those students whose results can be used in a fixed effects regression. This subgroup of students participated in a total number of 96 tests (the 'usable sample').

Figure 1 below shows the frequency distribution of the media types consumed before the three test days within the usable sample. We see a fairly even distribution across Types A, B and C, which is good news – most students adhered to the video type assigned to them. Only two test results were preceded by the consumption of more than one media type – too small a group to make any inferences. Fortunately, contrary to the instructions given to the students, a sizable number of tests were completed without the prior consumption of any of the tutorials. This set of test results can serve as a benchmark to ascertain whether the media have any learning effects at all. If all students had watched or listened to one of the tutorials as instructed, we would only be able to check for the differential effects of the three media types; the question whether they have any use at all would have to remain unaddressed.

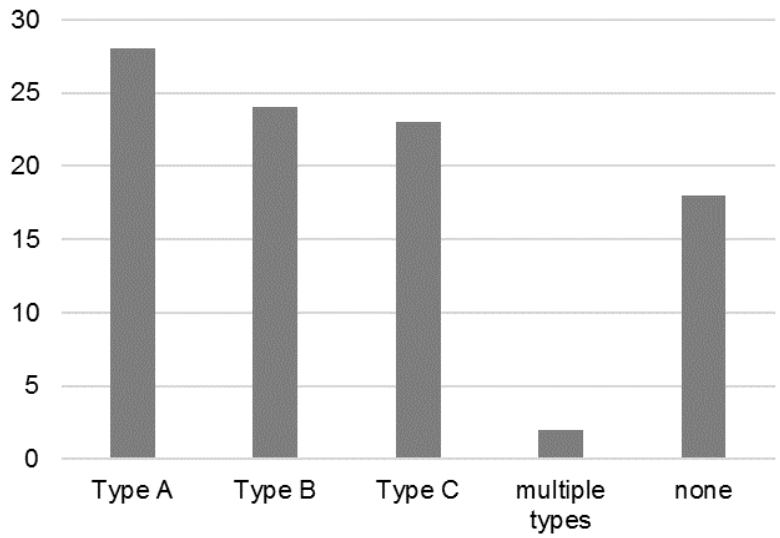

Figure 1: Media types consumed

As mentioned above, the students were also asked to indicate on which of the eight days preceding and including the test day they had consumed the media. This question was to serve a twofold interest: First, to simply observe the students' habits, and second, to test whether the time of consumption had any effect on performance – plausibly, content that was absorbed more recently will be better remembered in the test. Figure 2 below shows the resulting frequency distribution. Consumption more than one day before the test day occurred infrequently and is

therefore subsumed under one category. Evidently, most students consumed the media on the day immediately preceding the test. As for the second interest, it was sadly not possible to include this variable in the regression reported in the next section because the students' incomplete answers to this question would have overly curtailed the usable sample in a fixed effects specification. Nevertheless, inclusion of the variable in a random effects estimation of the full sample tentatively indicates that more recent consumption is indeed associated with better performance (results not reported).

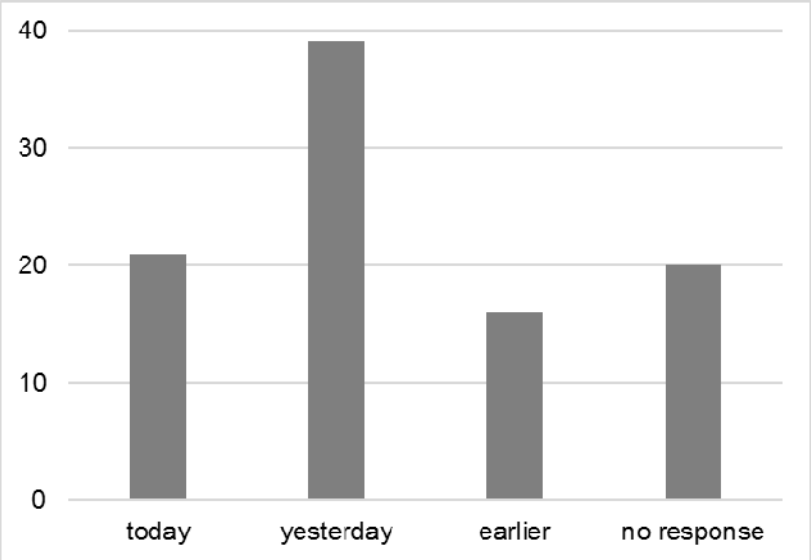

Figure 2: Day on which the media were consumed, relative to the test day

As a final piece of information before the lecture-related questions, the students were asked how many times they had consumed the media. The results were transformed into a dichotomous variable (repeated consumption – yes / no) because very rarely had students consumed a tutorial more than twice. Thus rephrased, 65 test results were preceded by one-time media consumption, 15 by repeated consumption, and no response was given in the remaining 16 cases. This variable was unsuitable for inclusion in a fixed-effects regression for the same reason as above. Yet, in the full sample, random effects estimation, in accordance with expectations, indicates that repeated media consumption indeed significantly improves the learning outcomes (detailed results again not reported).

3.2 DATA QUALITY

The data that was exported from the software which records the clicker responses unfortunately exhibited signs of poor quality. In a number of instances, the students evidently submitted incomplete and inaccurate information, either purposely (e.g. because they did not want to disclose their identity) or inadvertently (perhaps because they accidentally pressed the wrong clicker button or pressed the correct button but with insufficient determination). These conjectures about student behaviour are based on several observations:

- A number of students submitted fewer than the required five digits of their matriculation numbers; some even submitted no information at all that would allow them to be identified across the three test days. In some cases, such behaviour was likely deliberate. Other students, by contrast, participated in two or three test days and missed one or two digits of their matriculation number only on one of the days. Here, the omission was likely inadvertent. In cases where three or four digits were transmitted, it was sometimes

possible to deduct the full matriculation number and thus to match the specific test results with results produced by the same student on a different test day, increasing the usable sample size. However, this procedure is subject to a risk of error if the student in question participated in the test without having enrolled in the course.

- In 13 cases among the full sample, when asked about the media type consumed, students responded "none" even though, two questions further on, they did indicate a day on which their consumption had taken place, rather than not responding to this question. Similar inconsistencies exist between the questions on media type and repetition.
- In additional questions (which we turn to in the next section) on their personal experience with the media, some students transmitted a relative evaluation of the different types even though they simultaneously reported having consumed none of them.

The consequence of incomplete information is a reduction in the usable sample size; the consequence of inaccurate information is the introduction of random noise in the data, which will tend to inflate standard errors in the regression and thus to reduce the probability of finding significant effects. This we must bear in mind when interpreting the estimation results in the next section. Furthermore, the issue of data quality presents a major task for future research.

In a second iteration of this experiment, the questions and answer choices would also need to receive some revision. In one question posed on the first test day, the correct alternative from three given response options received 84% of the votes, with all of the remaining votes going to the second option, and none to the third. This question clearly made only a limited contribution to generating information about the students' performance. Furthermore, a minor mishap on the third test day also practically rendered one of the five questions obsolete: One of the five response options was accidentally formatted slightly differently from the other four, which virtually all students (correctly) interpreted as indication of this being the correct answer.

4 RESULTS

4.1 PERFORMANCE EFFECTS OF MEDIA CONSUMPTION

The main results pertaining to the performance effects of the different audio-visual media are summarised in Table 1 below. The dependent variable – the number of correct responses by a given student on a given test day – was regressed on a set of dummy variables for the different video types (with "none" being the omitted category), on dummies for the second and third test day, and on a set of dummies for the 40 students in the usable sample. Within this fixed effects approach, Poisson estimation was selected due to the count nature of the dependent variable. The coefficient estimates of the student dummies are not reported here, though a number of them are highly significant. The variables relating to the time of media consumption and repetition are not included in the model for the reasons given above.

Variable	Coefficient estimate	P > z
video type A	0.5339	0.020
video type B	0.6338	0.007
video type C	0.3726	0.163
multiple v-types	1.1984	0.003

test day 2	-0.3808	0.002
test day 3	-0.0145	0.891
Conditional Poisson regression with fixed effects at the student level. Robust standard errors. 96 observations from 40 students.		

Table 1:
Regression results.
Dependent variable: number of correct responses by student i on test day t

The results show that, relative to the first test day, the students performed significantly worse on the second day but only marginally worse on the third. This pattern may well be ascribable to the above-mentioned unintentionally simple questions on the first and third day, respectively.

Our primary attention, however, is directed at the performance effects of the different types of media. All three formats appear to promote the performance of their consumers relative to the case of no media consumption. Types A and B exhibit a significantly positive performance effect, Type B even strongly so. As for Type C, we cannot reject the hypothesis that its consumption makes no difference to performance. The estimate relating to multiple video types shall receive no further mention for the dual reason that this particular subsample comprises only two test results which, moreover, benefited from greater exposure to virtual teaching than those that were preceded by only one type of media. As a final lesson from Table 1, we note the ranking of coefficient sizes among the three media types: B>A>C.

To establish whether there is a significant difference in the performance-enhancing effect of the three media types, we run pair-wise X^2-tests for the equality of the respective regression coefficients. The results are reported in terms of p-values in Table 2 below. The p-values express the probability of obtaining, due to random sampling, a pair of coefficients with at least the given differential if in the 'true' relationship that underlies the data the coefficients are in fact equal. The only notable result is that Type B has a greater effect than Type C, but only if we apply the 10%-level of significance. We hypothesise that more clear-cut results could be obtained with better data.

Video Types	**B**	**C**
A	0.295	0.257
B	-	0.054

Table 2:
X^2-test of coefficient equivalence – probability values

4.2 STUDENT ATTITUDES TOWARDS MEDIA-ASSISTED LEARNING

At the end of the third test day, the students were asked to respond to two extra questions on their attitudes towards media-assisted learning. The first of these sought to elicit the degree to which the students agree with the statement, 'I appreciate the opportunity to prepare for lectures using audio-visual media'. Figure 3 below shows the frequency distribution of the responses. The data underlying Figures 3 and 4 cover all participants of the third test day, regardless of whether their results were included in the regression analysis.

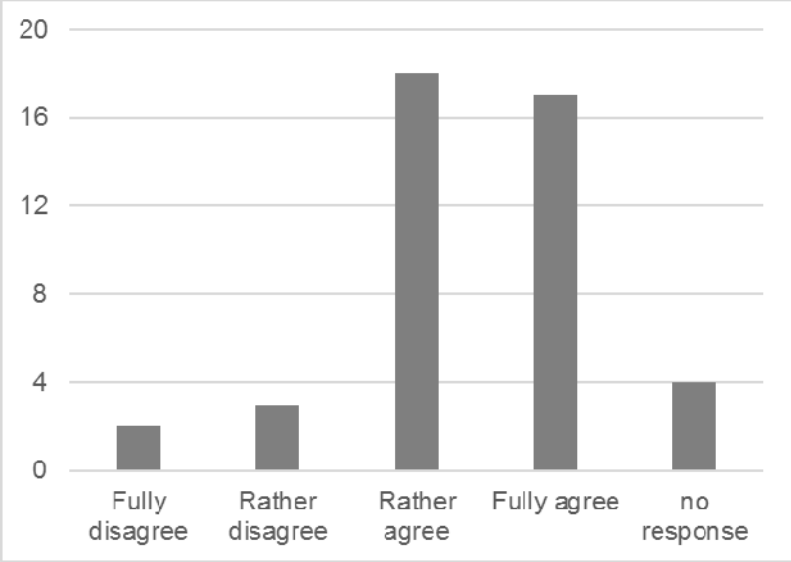

Figure 3: Responses to the statement: 'I appreciate the opportunity to prepare for lectures using audio-visual media.'

Clearly there is a great amount support for the use of such media on the part of the students. This comes as no surprise, considering the specific context of our experiment: The audio / video files were made available to the students not quite as an alternative, but more as a supplement to traditional lectures. Following the completion of each test, the contents were discussed in class so that the students were able to learn the relevant concepts even if they had not consumed the media. Hardly any cost was associated with them consuming or failing to consume the media; there was no trade-off. To them, the tutorials constituted an additional, optional source to learn from – which they predictably appreciated. The responses to this question will likely differ in other contexts, for example if students are presented with the choice of audio-visual tutorials as a strict alternative to lectures. Fischer and Spannagel (2012) list a number of ways in which video lectures can be used in relation to conventional lectures. In a setting not dissimilar to ours, the authors also find a strong preference among students for the use of electronic learning media. Furthermore, in any event the students' attitudes by themselves cannot guide the decision as to whether such media should be employed in teaching: In the evaluation, the students need not consider the substantial effort required to produce the media.

The second and final extra question posed on the third test day promises more interesting insights. The students were asked to indicate which of the three media formats they consider most beneficial to their learning progress. Figure 4 displays a count of the responses.

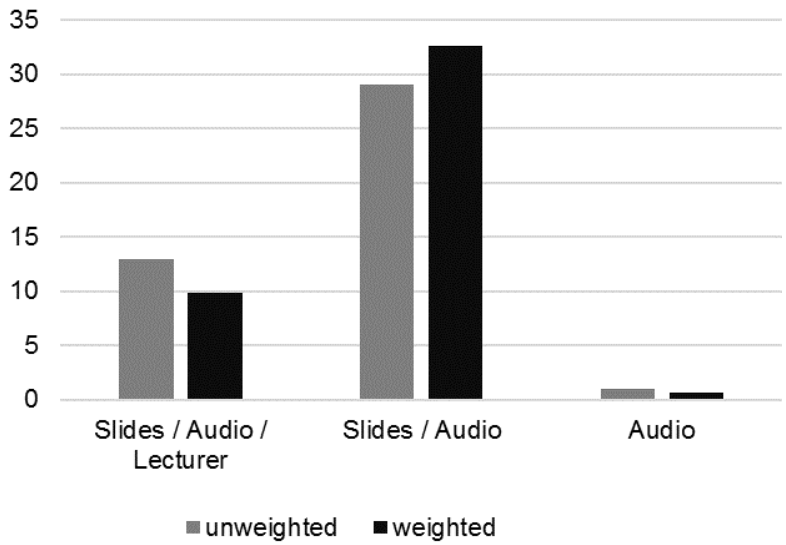

Figure 4: Responses to the question: 'Which type of media do you consider most beneficial to your learning progress?'

The bars shaded in grey represent the unweighted frequencies. They show a clear preference for media Type B, which receives more than twice as many votes as Type A, with Type C being preferred by only two respondents. The picture becomes even clearer if we weigh the responses by the number of media types that each respondent has consumed in total to take account of the fact that an evaluation by an experienced consumer should be more valuable than one submitted by an individual who has not consumed any of the tutorials. The ranking of the students' preferences thus matches the ranking that we obtained from their performance data.

Like the results obtained from the performance data, the students' preference for Type B is somewhat unexpected and stands in contrast to results obtained by Fey (2002) in a comparable setting. One might have thought that more information were always preferable, that the additional video component of Type A could only help the students' learning. Yet the results seem to remind us that the assumption of free disposal does not apply to information. Multi-media tutorials entail the risk of cognitive overload; due to our limited cognitive capacity, more information can actually obstruct learning (Sweller et al. 1998). Video imagery promotes learning in that - and only if - it serves to motivate and to elicit emotions in the learner (Edelmann 2000). Whether the videos used in this study were suited to achieving these objectives remains open to question.

5 CONCLUSIONS

Picture a lecturer who wishes to enhance her course with electronic learning media, perhaps even to 'flip' the classroom. She may wonder which of the ever-widening range of media formats to use. A survey of the literature will only be of limited value to her: The results are mixed, the methodologies are hardly comparable and the question remains to what extent recommendations derived from an experimental setting are transferable to her classroom.

In contrast to much of the literature, we have tested the learning effects of three different media formats in a standard lecture context, coping with the sporadic attendance and perfunctory participation of some of the students, amongst other real-life obstacles. A sufficient class size (50+), a set of clicker devices and some knowledge of statistics are the only prerequisites for the lecturer to replicate the experiment described here and thus to find out which format works best for her purposes.

We find that both voice-over slides and a split screen showing the slides and a video of the lecturer significantly improved the knowledge of a group of second-year management students. A mere audio recording of the presentation, by contrast, yielded no significant learning effects. Considering that the audio files required the same effort to produce as the voice-over slides, little

is to be said in favour of offering the former format to the students, especially since virtually none of them indicated a preference for the audio files.

Perhaps somewhat surprisingly, the voice-over slides marginally outperformed the videos, both in terms of learning performance and the students' preferences. This result seems to suggest that learning videos must be very well done, perhaps including animation and other more elaborate elements, for the risk of cognitive overload to be outweighed by their positive effects, including additional motivation and emotional involvement. Yet we agree with Fey (2002) in cautioning that any limited value-added of the videos' learning effects may not be worth their substantial production costs. We also caution not to overly generalise our - or anybody's - results because of the large differences between individual learning media even within a given format: No two lecture videos are alike and can be expected to have the same effect. This is all the more reason to advocate an ad hoc, 'on-site' investigation as to which format works best in a specific teaching context.

If this experiment were to be repeated, we recommend dropping the audio-only media as the dominated choice to focus on the differential effects of Types A and B. Also, efforts should be undertaken to obtain a larger and better data set. Possible measures to that effect include, for example, advertising cash prizes for students who perform well on the test and who submit all required answers to the auxiliary questions.

REFERENCES

Caldwell, J.E. (2007). Clickers in the large classroom: Current research and best-practice tips. Life Sciences Education, 6(1), 9-20.

Chen, C.-M., Wu, C.-H. (2015). Effects different video lecture types on sustained attention, emotion, cognitive load, and learning performance. Computers & Education, 80, 108-121.

Edelmann, W. (2000). Lernpsychologie. Weinheim: BeltzPVU.

Fey, A. (2002). Audio vs. Video: Hilft Sehen beim Lernen? Vergleich zwischen einer audiovisuellen und auditiven virtuellen Vorlesung. Unterrichtswissenschaft, 30(4), 331-338.

Fischer, M., Spannagel, C. (2012). Lernen mit Vorlesungsvideos in der umgedrehten Mathematikvorlesung. In: Desel, J., Haake, J. M. and Spannagel, C. (ed.), DeLFI 2012 – Die 10. e-Learning Fachtagung Informatik. Hagen/Heidelberg: Gesellschaft für Informatik, 225-236.

Hamburger Bürgerschaft (2015). Digitales Lehren und Lernen an den staatlichen Hamburger Hochschulen. Mitteilung des Senats an die Bürgerschaft. Drucksache 20/14262

Karnad, A. (2013). Student use of recorded lectures: a report reviewing recent research into the use of lecture capture technology in higher education, and its impact on teaching methods and attendance. London School of Economics and Political Science.

Kay, R.H., LeSage, A. (2009). Examining the benefits and challenges of using audience response systems: A review of the literature. Computers & Education, 53, 819-827.

Kundisch, D. et al. (2013). Classroom Response Systems. Informatik Spektrum, 36(2), 389-393.

Schmucker, S. (2015). Cognitive activation in mass lectures through electronic voting systems in the lecture theatre. In: International Conference on Quality in Higher Education - ICQH 2014, Proceedings Book, Sakarya/Turkey, 67-72.

Schmucker, S., Häseler, S. (2015). Measuring the impact of student diversity on performance with classroom response systems. The Turkish Online Journal of Educational Technology (forthcoming).

Simpson, V., Oliver, M. (2007). Electronic voting systems for lectures then and now: A comparison of research and practice. Australasian Journal of Educational Technology, 23(2), 187–208.

Sweller, J., van Merrienboer, J.J.G., Paas, F.G.W.C. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10(3), 251-296.

