UNIVERSITY OF
Southampton

University of Southampton

Faculty of Environmental and Life Sciences

School of Psychology

Notetaking in the Lecture Theatre – Examining the Impacts of Popular Encoding Strategies

by

Katie Alice Coria

Thesis for the degree of Doctor of Philosophy

December 2018

University of Southampton

Faculty of Environmental and Life Sciences

School of Psychology

Thesis for the degree of Doctor of Philosophy

Notetaking in the Lecture Theatre – Examining the Impacts of Popular Encoding Strategies

by

Katie Alice Coria

Table of Contents

List of Accompanying Materials

Research Thesis: Declaration of Authorship

Print name:	Katie Alice Coria

Title of thesis:	Notetaking in the Lecture Theatre – Examining the Impacts of Popular Encoding Strategies

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signature:		Date:	21/12/18

Acknowledgements

Firstly, I'd like to thank my external examiner, Kou Muryama and my internal examiner, Ed Redhead for taking the time to read my thesis and provide comments – I thoroughly enjoyed discussing it with you.

Secondly, I'd like to thank my supervisor, Philip Higham, who has provided invaluable academic guidance for the past four years, and to the University of Southampton for providing me with a Jubilee Scholarship to allow me to pursue this research. I would also like to extend this thanks to my participants – both current and former undergraduate students at the University of Southampton.

There were many times throughout the past four years when I hit (desirable?) difficulties with my research, and I am very grateful to my colleagues and friends: Ella Chadwick, Amy Boyson, Karlien Paas and Kenny Beeston-Brackstone for all of the support and friendship but most of all to my best friend, Lauren Griffiths, who kept me smiling throughout - your friendship is my greatest reward for this research.

To my husband, Oscar Coria, and my mother, Judi Henson, I offer my wholehearted thanks for the love, kindness and understanding when I have been stressed and less than pleasant. I also thank my dog, Coco, for sitting by me throughout, and offering me a paw when I look distressed. As a late addition, I'd like to thank my son, Joseph Coria, for being the greatest motivation in my life.

Finally, I would like to dedicate this thesis, with all of my thanks, to my father, Dr. John Henson, who encouraged me to pursue a Ph.D, and encouraged me to persevere. Your commitment and determination will always inspire me, and I hope I have made you proud in completing this research.

Literature Review

Since the beginnings of university lecturing, it has been considered good practice for students viewing the lecture to take notes on its content. The original, most obvious reason for this is to document the ideas, theories and principles discussed during the lecture for later review and revision prior to testing. Kiewra (1985) defined this as the *storage function* of notetaking.

However, with the introduction of computer software (such as Microsoft PowerPoint) into the lecture theatre in the 1990s, students were able to receive printouts of lecture content which they could subsequently review and revise at their leisure. This innovation would appear to negate the need for traditional notetaking if its sole purpose were that of content storage. Why waste the energy hand-writing notes when the computer can produce a neater, well-formatted and potentially more correct set of notes ahead of time? What this approach does not consider, though, is the process of encoding, or the *encoding function* (Kierwa, 1985). Encoding is the process by which the student initially records the information in their memory during the lecture and is considered to be the information retained for later recall even in the absence of review (Einstein, Morris & Smith, 1985).

Whilst notetaking is an active process that involves many techniques known to enhance cognitive processing such as retrieval practice, production and generation, the mere presence of a lecture slide handout changes this process. Longhand notes need to “stand alone” (i.e., make sense without supplementary material), so are usually written in full sentences whilst annotations on lecture slide handouts tend to be briefer and are often short sentences, words, or highlighting or underlining. Research has shown that students write significantly less content when annotating a lecture slide handout compared to making longhand notes (Marsh & Sink, 2010).

It is generally accepted that it is beneficial to engage in the techniques described above during learning and, it seems clear that, at least at the point of encoding, it is advantageous to engage with the process of notetaking. Does reliance on printed lecture slides mean that current students are learning less during a lecture than a student from their parents' generation? Considering the increases in tuition fees for tertiary study in the UK over the past decade, it appears to be more important than ever that learning opportunities are optimised for students to ensure that the learned material is retained during their undergraduate years and beyond to improve their employability and justify the cost of extra years in education. With this in mind, should lecturers be encouraging students to leave the slides at home? What is the impact of notetaking during a lecture on memory during later testing?

The Integration of PowerPoint into the Lecture Theatre

Microsoft PowerPoint (and it's competitors, such as Apple's Keynote) was initially developed to improve learning by making presentations more structured (Amare, 2000) and visually interesting to audiences (Harknett & Cobaine, 1997). By the mid 1990s, this software had begun to take hold in university settings, replacing the previously popular overhead projectors. Whilst their uptake varied between universities and subject areas, early research (such as Perry & Perry, 1998) indicated a positive reception of the new technology from students and lecturers alike. Aside from the obvious advantages of their material being reusable in future years and securely stored, the positive feedback from students encouraged lecturers to rely more on this method of presentation. As students are the "primary customers" of a university (Crawford, 1991), there is pressure to improve their satisfaction levels. Findings such as that of Harknett and Cobaine, who found that 80% of students surveyed felt that PowerPoint presentations benefitted their learning and Holzl (1997) who concluded that PowerPoint makes a lecture more attention-capturing and more visually clear for students. The positive early responses from students have hastened the transition and, by the early 2000s, PowerPoint had become commonplace in the lecture theatre.

Furthermore, Evans (1998), found that over the course of a semester, a group who received lectures through PowerPoint presentations not only reported reduced absences and less boredom but also scored higher on a final test than the control group whose lectures were delivered by overhead transparencies (mean scores of 81.6% vs 76.9% respectively). These findings were supported by Lowry (1999) who, in a similar between-subjects design found that current cohorts who received their lectures by PowerPoint scored higher on average in end-of-year tests than the previous cohorts, who had their lectures delivered through the use of overheads.

Although these findings indicate an entirely positive reception of this relatively new technology, not all studies found indications of improved academic performance. Szabo and Hastings (2000) conducted the first literature review into PowerPoint as a lecture delivery method before performing multiple studies looking at students' opinions, attitudes and performance scores. Whilst the results, like that of previous studies, found of memory recall for material given either using an overhead projector or a PowerPoint presentation during lectures, showed no differences between recall of material delivered. This indicates that the presentation delivery method alone was not sufficient to affect academic performance and the authors concluded that PowerPoint was providing more entertainment than learning enhancement and should be viewed as an auxiliary medium, alongside complimentary materials to improve educational value. Further to that, Bartsch and Cobern (2003) found that, although lectures featuring PowerPoint slides were preferred, students' performance was worse for quizzes that contained images or sound effects. However, it us unclear as to whether any accompanying printed slide handouts were provided alongside any of the PowerPoint presentations included in the above research, and any impact this could have had at either the encoding or revision period.

The (Perceived) Importance of the Lecture Slide Handout

Following these advances in technology and, as printing became more cost efficient, a complimentary material that quickly became popular was the lecture slide handout. These handouts traditionally consist of the content of the PowerPoint presentation scaled down so that multiple slides fit on A4 paper, for students to annotate during the lecture. This was predictably popular with students as, firstly, it offered them take-home content that was reliable and accurate and secondly, it removed their need to take longhand notes and replaced it with the option to annotate the handout. Frey and Birnbaum (2003) found that, as with the introduction of PowerPoint lectures, these new slide handouts had been almost entirely positively received by students. In an extensive questionnaire, the statement "PowerPoint handouts help me take better notes" had almost the highest agreement of any item, scoring a mean likert value of 4.1/5 although it is unclear what is meant by "helped"- whether the printouts are used in substitution for notetaking or for annotation with subsidiary notes. An indication of how lecture slide handouts may be used was that eight out of 50 participants in this research used a free-recall part of the questionnaire to volunteer a phrase similar in meaning to "I print out notes beforehand so I don't have to take notes". This implies that, at least for some students, the lecture slide handout was seen as a suitable substitute for note taking.

Further to this, Douglas, Douglas and Barnes (2006) found that when asked to rank all academic aspects of university life in order of importance, 'supplementary lecture materials' such as lecture slide handouts, were considered the fifth most important aspect in a list of more than 50 items.

Whilst the positive responses from students in these studies suggest that PowerPoint slides will increase learning, performance was not measured. A common assumption is that students know when they are learning, and can accurately predict their performance based on their learning experience. This, however, is known to be frequently incorrect as multiple studies investigating various judgements of learning and confidence ratings have shown that students are generally poor at regulating their accuracy in line with their confidence. Thus, they make inaccurate predictions of performance (Siedlecka, Paulewicz & Wierzschon, 2016; Matvey, Dunloksy & Schwartz, 2006) and tend to base their judgements on how easy they felt the material was to learn (Kelley & Lindsay, 1993). This is a potential fallacy that will be revisited later.

The Benefits of Notetaking vs Lecture Slide Handout Annotation – Current Evidence

Relatively limited research has been conducted into test performance following notetaking or lecture slide handout annotation. However, prior to the integration of PowerPoint and lecture slide handouts to the lecture theatre, a wealth of research demonstrated the benefits of notetaking.

A meta-analysis by Kiewra (1985) found that 35/61 studies reviewed found facilitative encoding effects for notetakers compared to those who simply listened to the information. Although this effect seems small, Cook and Mayer (1983) proposed that the effect would have been greater if the material used had been more appropriate for notetaking as many early studies used very dense lecture information or fast paced presentations, which were too rapid for note takers to follow. Equally, Peper and Mayer (1978) proposed that notetaking is an active process which helps learners to build connections between prior content and external knowledge – thus performance tests should measure generative learning (learning built from prior knowledge) rather than simple recall. Peper and Mayer concluded that the effect seen in a classroom would be greater than the effect measured in a laboratory setting as generative learning occurs when a student attends a series of lectures (such as those in an undergraduate module) and tests are designed to identify those who are able to integrate multiple concepts more than those used in a laboratory which typically measure single dimensioned recall based on a single encoding episode.

Furthermore, Einstein, Morris and Smith (1985) found that whilst students recalled equal numbers of high and low importance facts regardless of whether they took notes, those in the notetaking group recalled a significantly greater number of high-importance facts than those who observed a presentation without notetaking. This indicates that notetaking helps students to distinguish between important and non-important information during lectures.

Although the above findings indicate the efficacy of longhand note taking, all the research was conducted prior to the introduction of PowerPoint lectures and handouts, so it is unknown whether the presence of a lecture slide handout would have an effect on quantity or quality of information recalled when compared to longhand note taking. However, a study by Kiewra (1988) provides an indication. Participants were assigned to a study group in which they were either given a complete handout of printed notes to accompany the lecture (similar in content to a lecture slide handout but not exactly matching the presented lecture content), a skeleton printout (a handout containing some basic information with space for the participant to elaborate) or blank space. Given that there was no revision session (thus no retrieval practice), the quantity of information recalled by participants could be attributed to the method of encoding. Results showed that the groups provided with a blank space or skeleton printout produced higher recall of information than those who were given a full-text handout. Although the full-text handout differs from a modern slide handout in format and text quantity, it served a similar function in that it removed the necessity to write any more longhand notes than simple annotations. Therefore, this early research into notetaking implies that the process of taking longhand notes differs in the presence of handouts, and that the action of taking notes facilitates encoding and promotes greater long-term recall.

Findings indicative of the efficacy of notetaking have persisted into the 21st century with Kobayashi (2005) examining variables such as schooling level, presentation mode, and type of learning outcome on final recall when notetaking is compared to non notetaking (which, in almost all cases was simple lecture observation). The findings of this meta-analysis show a positive but modest effect of notetaking, regardless of other variables, but that presentation type can interfere with the notetaking process if presentations are delivered in a rapid or complex manner. Neither schooling level nor learning outcome affected notetaking, implying that this modest benefit of notetaking subsists across all age categories, academic levels and topic areas. However, as with previous research, this meta-analysis did not focus on comparing longhand notetaking to slide annotation and therefore it is unknown how schooling level, learning outcomes or presentation mode could affect encoding when students were annotating a lecture slide handout. Despite this, Kobayashi (2005) also concluded that the effect of notetaking was larger for free recall conditions than for recognition tests. In terms of university exams, this research indicates that notetaking would be more likely to improve a students' performance for free recall (such as essays or extended answer questions) than for recognition tests (most commonly, multiple choice). This is one aspect that was described in the above paper as a potential contributory factor to the modest effect size, as the majority of tests investigating recall (over 60%) used testing strategies other than free recall with most using multiple choice.

Although Kobayashi's research outlines many advantages of longhand notetaking, the lack of comparison to lecture slide handout annotation means we cannot draw conclusions about the efficacy of notetaking vs slide annotation based on this research. However, some research directly comparing the two strategies has shown inconclusive findings.

Murphy and Cross (2002) investigated the impact of lecture slide handouts by asking biology undergraduate students to self report if they consistently use slide handouts within their lectures and comparing this to their end-of-year scores with those who reported that they do not use slide handouts in lectures. Results showed that, over three years of cohorts, participants who reported not using slide handouts scored significantly higher in end-of-year tests. This further supports the above findings with regards to the efficacy of longhand notetaking but it is important to consider that this research was undertaken for educational purposes and without theoretical basis. As such, there was limited experimental control and elements of bias. Firstly, the ambiguous use of the term "consistency" when referring to frequency of slide handout use could have led to discrepancies in interpretation between participants. Secondly, the natural setting of the research led to the occurrence of a bias; this research reported that the majority of participants (over 80%) who claimed that they did not use lecture slide handouts were in their second and third year of undergraduate study. These participants would be more acclimatised to the demands of university study and more likely to score higher. Equally, 86% of the students who reported not using lecture slide handouts had studied biology at high school, compared to 74% of those who reported using lecture slide handouts and thus could have had an advantage academically. Whilst these results indicate an advantage to making notes over slide handout annotation, it also demonstrates a potential metacognitive understanding in a more naturally academic student, who would recognise that writing their own notes is beneficial for their learning.

Following this research, Weatherly, Grabe and Arthur (2003) conducted a similar study but without the bias of the self-report discussed above. In this research there were two classes – one with no access to lecture slide handouts before the class and one with access to lecture slide handouts prior to the class. As with the above research, results showed that those who were required to make longhand notes “outperformed” those who had access to the lecture slide handout in advance of the lecture. Although this provides further support for the advantage of note taking, the natural setting of the environment meant a lack of control thus it is unknown whether performance relates to encoding or later revision. This lack of control relating to data collection in the natural setting of a lecture will almost inevitably lead to a lack of distinction between any benefits that occurred at encoding and later restudy, when examining test performance.

Similarly, Babb and Ross (2009) used a between-subjects naturalistic design, and manipulated lecture handout availability over two semesters. The conditions were counterbalanced; both courses within the study were run both in the first and second semester of the academic year, and participants studied both courses. However, for one course, participants received a slide handout before the lecture, and in the other, they were not provided with a slide handout until after the lecture, thus assuming participants would make longhand notes. Unlike Murphy and Cross (2002), Babb and Ross found that class attendance improved when slides were provided in advance, but there was no difference in exam performance, which implies no differences in encoding regardless of lecture slide handout presence. However, two confounds make it difficult to draw firm conclusions from this study. Firstly, the only measurable variable that was manipulated was whether the handouts were present prior to the lecture. Given that the experiment was part of a real university course, students themselves made decisions as to whether to download and print slide handouts in the condition when they were available before the lecture – indeed, 18% stated that even when slides were available, they did not download them until after the lecture and a further 19% said that they never downloaded the slides. With such a high proportion of the class not adhering to the experimental conditions of the research, it is difficult to draw firm conclusions. Secondly, and more importantly, due to the naturalistic design, it is not possible to isolate the effect of encoding on later testing when many factors, particularly restudy between encoding and testing are likely to have a significant influence on the outcome.

The criticisms for the above studies demonstrate that it is difficult to adequately control for extraneous variables when collecting data in a natural setting (in these cases, in an undergraduate lecture). This is combined with the obvious disadvantage that it is impossible to separate the storage and encoding functions discussed above (as revision cannot be adequately controlled or manipulated) when considering factors contributing to student performance in the testing phases. However, this research has provided sufficient indication that this is an area worthy of further research in a more controlled environment.

Marsh and Sink (2010) conducted the most experimentally controlled research investigating the effect of notetaking vs slide annotation on encoding. Prior to their experiments, they conducted a brief survey into notetaking habits amongst university students, in which 74% of their students stated that they preferred to receive a handout prior to the lectures. This is consistent with other literature on student preference discussed previously (e.g.,Douglas, Douglas & Barnes, 2006; Frey & Birnbaum, 2003). In two studies, they experimentally manipulated the way in which participants took notes in a pre-recorded lecture. In two lectures, participants were provided with both blank paper and a lecture slide handout (the “handout-present” conditions), and in a further two lectures only blank paper on which to make longhand notes (the “handout-absent” condition). Two of the lectures (one from each of the “handout-present” and “handout absent” conditions) were then reviewed prior to testing and the other two were then tested with the absence of review, in an attempt to isolate the encoding effect of both conditions. In both experiments, Marsh and Sink found that for both immediate and delayed testing, participants performed equally across both conditions in the absence of restudy. Furthermore results showed that, following restudy, both when they were allowed unlimited review time (Experiments 1 and 2) and when they were given a set time of 2-min (Experiment 2), participants scored better in tests corresponding to the lectures viewed in the “handout-present” conditions. Marsh and Sink concluded that by providing a lecture slide handout to accompany the lecture, students can dedicate their entire attention to the lecture itself, thus reducing the need for this dual attention, which should increase memory recall on a later test. Marsh and Sink named this ability to focus exclusively on the lecture “efficient encoding” (p.692), as a contrast to Bjork’s (1974) concept of Desirable Difficulty, which in this circumstance refers to the idea that making notes whilst attending to the lecture provide a difficulty which, when overcome, will provide greater memory at later testing. This concept will be returned to later in this chapter.

Whilst this research appears to support the distribution of slide handouts prior to a lecture, two aspects of the methodology affect the validity of this claim. Firstly, in the “handout-present” lectures, participants were also provided with blank paper, with which to make longhand notes. An analysis of note content within this experiment revealed that these participants did use the blank paper, although they wrote half as many notes on average. This leaves the effect of the handout inconclusive, as there is no way of deciphering which method (using the handout or making longhand notes) was actually used by the participants in the “handout present” condition. Indeed, it is entirely possible that under this condition, some participants paid little attention to the handout, but acknowledged its awareness for later revision, thus made fewer notes, but still noted down the key material, so were able to perform equally with the “handout-absent” group. Secondly, there was no control, or “reference” condition in which participants did not make notes. Hence it is not clear from these results whether the equal test performance between these groups was due to both handouts and notetaking being ineffectual, or whether each had beneficial effects specific to one group only.

Although the multiple studies used in this research were robust and easy to replicate, there have been no further findings to support this research. The majority of findings in this area, although still limited, appear to indicate contradictory findings.

The Cognitive Processes Involved in Notetaking vs Slide Annotation

Kobayashi (2005), Marsh and Sink (2010) and other research (such as Schmalhofer, McDaniel & Keefe, 2002, and Bui, Myerson & Hale, 2013) agreed that it is not just the act of taking notes but how these notes are taken which can affect performance. It is known that a variety of practices occur during notetaking, some of which are considered more effective than others. To be able to predict the effect of either longhand notetaking or annotating a lecture slide handout, it is necessary to examine the processes involved in both strategies during a lecture.

Firstly, it is necessary to make the assumption that when annotating a lecture slide handout, a student does not replicate the material on the printout in their own writing but merely makes extra annotations. This assumption is supported by surveys into student handout use such as Clark (2008). In this case, the material is already written thus the student is not involved in the process of production of information or the generation of further information (two cognitive processes discussed below) but is more likely to be attending to the lecture entirely, or reading the handout alongside the lecture, occasionally making extra, short notes or highlighting passages. Whilst it is possible that students are using the handout to make longhand notes in their own words, research, such as Marsh & Sink (2010) who found that students who annotated a handout wrote half as many words as those who only wrote longhand notes, suggests students do not tend to practice this.

The three strategies mentioned above (observing the lecture, reading the handout alongside the lecture and highlighting the handout) are passive strategies, which are considered to be inferior for encoding than active strategies, in which the student is forced to become involved in the process of learning (Einstein, Morris & Smith, 1985).

Some students may take a handout for purposes of later revision and choose to simply observe the lecture as it is given. This process is entirely passive. Equally, a student may elect to read the handout as they observe the lecture. Reading the content of the handout in this way alongside the lecture provides a slight advantage, as the student is engaging in the process of recognition (Bromage & Mayer, 1986). According to the quantitative hypothesis (Mayer, 1983), rereading increases the amount of information encoded as information is attended to for a second time. However, there have been mixed findings regarding the efficacy of rereading as an encoding strategy and some findings (e.g., Verkoeijen, Rikers & Ozsoy, 2008) indicate that an important factor in the efficacy of this strategy is the spacing between initial reading and rereading. This research found that increased time spacing between initial reading and rereading improves recall whereas in a lecture, the time between reading and rereading is likely to be negligible if the student wishes to keep up with the lecturer's pace of delivery.

Highlighting passages of the lecture slide handout during lectures is another common method used, which has been found to be popular amongst students (Wollen, Cone, Margres & Wollen, 1985). There has been some support for the use of highlighting when encoding information, such as the isolation effect (Hunt, 1995), which claims that highlighted items are semantically distinct and unique so potentially better remembered. However, this is only beneficial if students are discerning about what is highlighted. Blanchard and Mikkleson (1987) found that active highlighting (where the information is actively and correctly selected within the context of the text, which involves the reader making decisions about what is important) showed a slight advantage when recalling the highlighted information over those who just read the text. This though was offset against poorer recall in the highlighting condition for information not highlighted. This suggests that highlighting material during a lecture may not be an optimal strategy for recall, especially as a student is often being exposed to novel and unfamiliar material, thus may not be able to discern that which is important to highlight. Dunlosky, Rawson, Marsh, Nathan and Willingham (2013) concluded in their meta-analysis of study strategies that highlighting is a “low utility learning technique” (p.45).

Conversely, in the absence of a lecture slide handout a student can select either to passively observe the lecture (discussed above) or to make longhand notes to record the lecture information. The number of techniques which they could elect to use when making these longhand notes is virtually unlimited, and the benefits associated with these strategies are summarised in Table 1.

Regardless of how the information is presented, a notetaking student will almost inevitably write some of the content in verbatim form, from either the lecture presentation or the words of the lecturer and thus the student is reproducing the material. This reproduction could be seen as semantically similar to repeating the information out loud, which has been found to have a substantial benefit to memory recall when compared to silent reading (MacLeod, Gopie, Hourihan, Neary & Ozubko, 2010). This is known as the production effect, which has been studied in depth and is considered to be a robust method for improving recall. However, one of the main reasons why it is considered to be successful is because repeating a word out loud makes it become distinctive in comparison to the words not read aloud (Dodson & Schacter, 2001). MacLeod et al. (2010) supported this theory, finding that the production effect only occurs within subjects and not between. This appears to imply that the production effect would have limited value in a lecture as only the words written down would be remembered, and, as discussed above, students exposed to unfamiliar materials may not be able to accurately identify the important aspects necessary to be recorded. Despite this, Ozbuko, Hourihan and MacLeod (2012), have found that the production effect endures beyond short-term reproduction and appears to lead to deeper understanding of the produced material. The authors postulated that reproducing information caused participants to process this information more deeply, leading to improved understanding. If correct, this would indicate that by producing the information by re writing it, note takers would have improved recall and better understanding of the material than those given the information to read during the lecture. Indeed, Mama and Icht (2016) found an even greater advantage for writing information compared to vocalizing it. These results imply that copying slide information could benefit memory more than simply reading information on a slide handout.

A recent study by Bui, Myerson, and Hale (2013) supports this conclusion. They found that when participants reproduced as much of the verbal content given during the lecture as possible (verbatim notetaking) this led to very good performance on an immediate test. In fact, it was even better than the performance of a group of students instructed to paraphrase and organize their notes. This advantage was short lived however, as it was not just attenuated, but *reversed,* on a delayed test taken 24-hours later. Nonetheless, Bui, Myerson and Hale's results suggest that verbatim notetaking can confer some memorial advantages, at least in the short term. Additionally, verbatim notetaking may have other advantages. For example, it may reduce the demands on attentional resources that derive from more generative notetaking processes such selecting and paraphrasing important information (Cohn, Cohn, & Bradley, 1995). However, over a longer period, such as between a early-term lecture and an end-of-term exam, the results of Bui et al.'s results imply a more generative strategy of notetaking is likely to hold further benefits compared to production.

This strategy of paraphrasing the information or rephrasing it so that it is better understood is otherwise known as self-generation. The benefits of self-generation for learning, or the generation effect, has had considerable support since it was first introduced by the findings of Slamecka and Graf (1978), who concluded that generating all or part of an item is better for memory than reading it. Initial research into this effect most frequently looked at word pairs, in which participants were asked to generate all or part of one of a word pair (for example, "Hot, C__d", or "What is the opposite of Hot?" vs reading "Hot, Cold"). Findings by Fielder, Lachnit, Fay and Krug (1992) suggested that the more generation that is required, the better it would be recalled. This implies that this effect should be found amongst longhand note takers in a lecture theatre, as they would be required to generate all of the information. A recent meta-analysis by Bertsch, Pesta, Wiscott and McDaniel (2007) summarized the findings of 86 generation effect studies and compared results looking at variables such as participant age, recall test type and experimental design. They found a subsisting, moderate to strong effect for the generation effect and proposed several theories as to why generating information appears to improve its likelihood of being recalled, one of which is that generation is a cognitively difficult strategy, something which supports the principal of desirable difficulty, and will be returned to later.

Mueller and Oppenheimer (2014) further supported this theory of the benefit of self-generation as a notetaking strategy in their study that investigated the use of computer vs pen-and-paper notetaking. They concluded that as the students using a pen and paper were not able to write as rapidly as those typing on a computer, they were increasingly reliant on paraphrasing and generating the information in their own words, and performed better at testing because of it than those typing, who tended to reproduce the content of the lecture verbatim.

Another potential reason why self-generation of information during notetaking could be beneficial is because of its similarity to self-testing. Whilst the time between receiving the information from the lecture and generating the rephrased content in note form is not long, students must store the information long enough to process its semantic meaning before either retrieving it to re-write or generating equivalent information in different words, both of which are a form of retrieval practice.

There is a vast literature demonstrating the benefits of retrieval practice between learning and testing (e.g., Karpicke & Blunt, 2011; McDaniel, Anderson, Derbish, & Morissette, 2007; Roediger & Butler, 2011; see Roediger & Karpicke, 2006 and Rowland, 2014 for reviews). Indeed, Dunlosky et al. (2013) considered it to be one of the highest utility learning techniques of those reviewed by them. Fazio, Agarwal, Marsh, and Roediger (2010) found that whilst confidence for final test answers was significantly lower following practice testing in an academic environment, accuracy was significantly higher, indicating that students are unaware of the value of practice testing on recall. A review by Rawson and Dunlosky (2012) found that practice testing is a reliable and efficient method of learning information, as those who are tested appear to be able to recall correct answers more rapidly than those who are exposed to relearning conditions, and the testing appears to induce less boredom. This research was supported by a meta-analysis performed by Rowland (2014) who further discussed some mechanisms behind the efficacy of the testing effect. Rowland concluded that there are several interacting properties contributing to this effect, such as semantic elaboration (the process of rehearsal and production of new words based on previous knowledge) and effortful processing (the difficulty in retrieving the information during practice testing makes it more memorable for later recall, Karpicke & Roediger, 2007). Recent research has even indicated that there could be a forward effect of testing, in that the retrieval practice of previously learned information could improve the learning of subsequent information (Pastötter and Bäuml, 2014).

Despite these benefits associated with testing, retrieval practice has mostly been discussed as a strategy that can be adopted after initial learning is complete (e.g., as an effective exam revision activity). However, some other research has found that it may be beneficial even during encoding. Di Vesta and Gray (1972) found that testing during an interval following a 5-min recorded communication improved later test performance compared to no testing, and Lyle and Crawford (2011) found that testing students immediately following a lecture produced significantly higher exam scores than those given time to review the content. Although the above research is more similar to a notetaking scenario, no research has been conducted exclusively investigating retrieval practice *during* a lecture as part of notetaking.

The efficacy of notetaking in the form of retrieval practice may depend in part on how accurate participants are at retrieving the correct information. If participants retrieve nothing but erroneous information (i.e., errors of commission), then retrieval practice could backfire. Specifically, this erroneous information could become strengthened in memory by retrieval practice, thereby increasing the likelihood that it is reproduced on later tests, worsening performance rather than enhancing it. Alternatively, there could be total retrieval failure (i.e., errors of omission), in which case the student would benefit less from this form of notetaking than if they had simply observed the lecture (cf. the bifurcation distribution model – Kornell, Bjork, & Garcia, 2011).

When comparing the processes involved in encoding when annotating a lecture slide handout versus when longhand notetaking, the research described above clearly demonstrates that the processes involved in notetaking are more conducive to later recall than those involved in lecture slide annotation. One undeniable similarity between all of the processes involved in longhand notetaking is that they all require considerably greater cognitive effort than those undertaken when a lecture slide handout is present. Piolat, Olive, and Kellogg (2005) stated that notetaking is a "complex activity that requires comprehension and selection of information and written production processes...the time urgency in selecting key points and recording them while comprehending new information at the same time places significant demands on the central executive and other components of the working memory" (p. 291). This appears to suggest that the difficulties associated with notetaking make it undesirable. However, is introducing obstacles to increase the demands on memory necessarily disadvantageous for learning?

Theoretical Explanations for the Benefits of Notetaking

A summary of these theories can be found in Table 1.

Desirable Difficulty. Some of the literature looking at the benefits of handouts discussed previously (e.g., Marsh & Sink, 2010) refers to ease of processing as a desirable aspect of learning. There is a common belief amongst students (and some lecturers) that, by improving the conditions surrounding learning (such as speaking very slowly and clearly and, in this case, providing lecture slide handouts so that the student can focus on the lecture), information will be more smoothly and fluently encoded and better recalled at later testing. However, a growing body of research applying principles from cognitive psychology to education has found the reverse to be true; generally speaking, it is *difficult* and *slow* learning that promotes long-term memory retention

The name given to this concept is *desirable difficulty*, and it was first described by Bjork (1994) as a condition in which obstacles are initially created for the learner that slow the rate of initial learning but once overcome lead to better retention and improved recall. Bjork concluded that the cognitive processes involved in understanding and overcoming the difficulties optimise long-term retention and transfer for the learner. A series of manipulations of learning conditions such as manipulating learning environment (Smith, Bjork & Glenberg, 1978), presenting information in a less structured format (Mannes & Kitnsch, 1987) and reducing perceptual fluency (Reder, 1987) all supported this concept of desirable difficulty, indicating that the unpredictability and disfluency associated with these variations led to improved recall. The cognitive demands involved in notetaking could be acting as desirable difficulties to promote encoding and maximise retention of information for later testing. However, it is important to acknowledge a limit to the benefit of increasing the difficulty of a task. Bjork and Bjork (2011) emphasise the importance of the word desirable, stating the necessity that the process optimises encoding and retrieval. If the difficulty of the task becomes unsurmountable, and if the learner is unable to overcome the obstacles necessary for the information to be processed, the difficulty becomes undesirable and inhibits the encoding process.

New Theory of Disuse. Bjork and Bjork (1992) proposed this theory as a modification of Thorndike's (1914) law of disuse. Thorndike's law states that memory traces decay over time if unused and are eventually lost. Bjork and Bjork's theory states that memory is more complex, and that there are other aspects to memory such as interference with other information, which can prevent memory from decaying, even over time, and that, as with desirable difficulty (Bjork, 1994 – discussed above), increased difficulty retrieving the memory trace can promote *improved* memory at later recall. This theory proposes two different types of memory strength: retrieval strength and storage strength.

The *retrieval strength* of a memory refers to how accessible it is at the given time. Its accessibility is dependant on many factors; such as how recently the information was stored and the context of retrieval. Retrieval strength ultimately determines whether a memory is recalled or not. Within the context of our research, retrieval strength would determine whether the student was able to answer a test question correctly.

The second component, the *storage strength* of a memory, is considered to be an index of learning, which reflects how well an item is related to other items in memory, and also increases with repeated study and retrieval. The storage strength of an item can only increase.

Although both storage strength and retrieval strength are separate, they interact with each other in order to remember an item. The increase in storage strength is caused by retrieval, but an item with *lower* retrieval strength (i.e. more difficult to remember) will cause an increase in storage strength when remembered. Therefore, the amount of learning caused will be greater if the retrieval strength is lower.

Retrieving an item increases both storage strength and retrieval strength in the short term. However, the increase in retrieval strength can be lost following interference or a time delay. This explains how items can be very easy to remember at one point (have high retrieval strength) but very difficult to remember at a later point (have low retrieval strength). However, the storage strength is unaffected by this time delay.

In terms of classroom learning strategies, it would be expected that notetaking and annotation of a handout would equally increase retrieval strength in the immediate term, but the need to re-word or condense the material amongst notetaking students would mean that they were reliant on using more difficult strategies, which could reduce retrieval strength in the short term. By delayed testing, storage strength reliance emerges and the difficult retrieval associated with the strategy of notetaking would increase storage strength more than the passive act of slide handout annotation, and these high levels of storage strength protect against forgetting, thus lead to improved performance at testing.

Levels of Processing Model. Craik and Lockhart (1972) proposed that the deeper that information is processed, the longer the memory trace will last. The entire concept behind this theory is that memories are entirely a result of how the information is processed. According to the theory, there are three different ways in which information can be processed: structural processing, phonemic processing and semantic processing. In the context of a lecture theatre, structural processing would relate to the physical appearance of the slides being delivered by the lecturer, and how they look, phonemic processing relates to sound so would be related to attending to how the lecturer is delivering the lecture, whereas semantic processing relates to the meaning of the lecture content. Craik and Lockhart proposed that semantic processing is the only form of deep processing, as it leads to meaningful analysis of the material and better recall at later testing.

When comparing longhand notetaking to lecture-handout annotation, it appears as though annotation lends itself to structural and phonemic processing, the shallower forms of processing within this model. This is because the student has the material present thus can afford to focus on how the lecture is delivered. However, in order to take longhand notes, a student will be more likely to semantically process the information in order to both a) decide what information to note down, as it is usually impossible to copy down the entire lecture verbatim and b) paraphrase the information into their own words. According to this theory, this would indicate that longhand notetaking would lead to deeper processing thus better retention of the information and greater quantities of recall at testing.

Integrating the Theory into a Framework

It is clear that the processes involved in longhand notetaking (discussed previously) make it more difficult than annotating a lecture slide handout, which is supported by numerous theories. Based on the above research, we predicted that students engaging in longhand notetaking will out-perform students annotating a lecture slide handout. This is what we investigated throughout this research. However, due to multiple factors affecting learning, and how learning is measured, we also examined the potential limits to the effectiveness of this strategy and whether longhand notetaking during lectures is, indeed, an optimal strategy.

Jenkins (1979) proposed a tetrahedral model of learning, in which he outlined four interlinking key factors of experiments that can be manipulated to affect learning outcomes. This model could be applied to learning in lectures, and we will attempt to use some aspects of this in our research.

The first factor in this model is the events that occur during learning. In the more general setting, this covers how the material is learned i.e. the type of learning environment. In the context of a lecture theatre this could mean the way in which the lecture is presented (whether the lecture is presented rapidly, or fluently, for example) or the material covered in the lecture (both the subject studied, e.g. geography vs medicine, and the content of the lecture studied, e.g. something fact-based or something relying on concept understanding).

The second factor in the model refers to the criterion task used to measure learning. At university, this learning is generally measured by end-of-unit tests, and studies conducted investigating learning in universities try to mimic this, although the delay period between learning and testing is frequently shorter than the typical delay seen in a real university scenario. Bui, Myerson and Hale, 2013, used a 24-hour delay before the final test, and Karpicke and Roediger (2008), used a week-long delay prior to testing in their experiments). The decay in memory for learned information over time has been accepted since Ebbinghaus (1885). However, no research has currently investigated whether this rate of decay varies between notetaking versus slide handout annotation for a time period longer than one week.

Furthermore, the way in which learning is tested also falls into this factor. The two ways researchers test memory are through recall (either free- or cued-recall), or recognition (for example, multiple choice tests). Kobayashi (2005, – see also Weener, 1974) found the benefits of notetaking to be greater for recall items than recognition tests. Therefore, it is possible that the method used in the end-of-term test could dictate the best methods for students to take notes throughout the course. Despite this, it is important to acknowledge that recall is considered to be more difficult than recognition, and that the above research only compared notetaking to passive observation. Thus, we cannot draw firm conclusions about whether any benefits of annotating a slide handout would be seen for either test type.

The third factor in the model is the subjects themselves. A vast range of interpersonal characteristics can affect learning, both in a university scenario and during a memory experiment. Various researchers have identified a number of characteristics that could restrict any benefits of notetaking, such as an individual's handwriting speed

(Peverley et al., 2013), attention deficits (Vekaria, 2011) and cognitive processing skills (Williams & Eggert, 2002). Equally, the motivation of a participant could affect engagement with the task, or how the task is interpreted. For example, research by Nestojko, Bui, Kornell, and Bjork (2014) showed that participants who expected to later teach the material that they were learning engaged in more effective learning strategies.

The final factor discussed by Jenkins is the learning activities undertaken which, in our case is the main variable that we intend to manipulate. The key activities that we intend to investigate are longhand notetaking versus slide annotation. However, there are various other techniques that have been researched (such as guided notes, see Markovits & Weinstein, 2018). Guided notes are handouts provided by the lecturer that give some structure but not all of the relevant information, leaving the student to generate the content. This activity of guided note annotation incorporates many of the more desirably difficult cognitive processes of longhand notetaking, but without the requirement that the student simultaneously attends to what the lecturer is saying whilst writing comprehensible notes. This act of notetaking whilst attending to the lecture is described as a “dual-task situation” (Craik, Govoni, Naveh-Benjamin & Anderson, 1996), and is considered to have negative consequences on memory (Mulligan, 1998).

Finally it is important to acknowledge that, although we will be focussing on learning at the point of encoding, students will inevitably restudy much of the lecture content prior to testing in a real-life university scenario. Previous research has almost unanimously concluded that reviewing the lecture content prior to testing improves test performance (e.g. Kiewra et al., 1991). As discussed previously, the information on a lecture slide handout contains a greater quantity and quality of information (as it is made ahead of time by an experienced lecturer rather than within the time limit of a lecture by a note-taker who is unfamiliar to the subject) than handwritten notes. However, we are not aware of any publications reporting on research investigating the impact of encoding condition on testing following restudy, where the longhand notetaking vs slide annotation groups are isolated. Research into other interventions aimed at improving encoding (such as spacing during learning, Rawson & Dunlosky, 2013, and practice tests Friedman, Muryama & Bjork, 2014) found that any benefits seen following encoding were attenuated by restudy.

Summary

Throughout this research, we test the theory that longhand notetaking during lectures is a more beneficial activity at the point of encoding than annotating a lecture-slide handout, whilst using the model to identify any mitigating factors or circumstances in which it ceases to be effective. We did this by comparing longhand notetaking and lecture slide handout annotation to a passive observation control group, whilst manipulating lecture material (across all lectures): lecture content (Paper 1, Experiment 1), lecture delivery (Paper 1, Experiments 1 & 2), length between encoding and test (Paper 3, experiments 1 & 2) and test type (all experiments). In addition, we integrated further encoding activities to test their efficiacy (Paper 2). Finally, we investigated any benefits of longhand notetaking following restudy (Paper 3, Experiment 2). We hope that through this research, we can help inform universities on the best policies to implement with regards to the provision of handouts prior to lectures, in order to improve the quality for students in Higher Education Institutions.

Table 1.

Summary of the Benefits of Longhand Notetaking

Theory/effect Name	Impact on Notetaking
Production Effect	If a notetaker writes down the exact content as seen/spoken in the lecture, reproducing tis information given in the lecture on to paper will promote retention of the information, at least in the short-term
Generation Effect	Paraphrasing or re-wording the information during notetaking is considered to be *more* beneficial still than reproducing the information verbatim, and has been shown to lead to greater retention of information in the longer term
Retrieval Practice/the Testing Effect	The time between the information being received and written down constitutes a form of retrieval practice. This retrieval practice is considered to be more effective at later testing than re-reading, which is similar to a strategy that a student would engage in when annotating a lecture handout. This advantage seen for retrieval practice is known as *the Testing Effect*
Desirable Difficulty	Longhand noteaking is comparatively more difficult than slide handout annotation. However, this theory suggests that the associated obstacles in longhand notetaking, once overcome, lead to improved retention of information compared to easily-learned material
New Theory of Disuse	Two different memory strength components: retrieval strength – how easily accessible a memory is, and storage strength – an index of learning, interact to determine whether information is remembered. An inverse relationship between retrieval strength and storage strength means that an item which is difficult to retrieve initially becomes more immune to later forgetting. Therefore, whilst longhand notetaking is considered more contextually difficult and would reduce initial retrieval strength compared to slide annotation, this would increase storage strength so the item would be more likely to be remembered at later testing.
Levels of Processing Model	How well the information is encoded leads to how well ti is remembered. Longhand notetaking generally leads to deeper processing of information. Namely, semantic processing where a student processes the meaning of the information, which leads to deeper encoding and better memory for the information when compared to slide-handout annotation, where a student is able to focus more on structural and phonetic processing, which is considered shallower forms of

	processing and less likely to lead to retention of the information.

Paper 1 Overview

With the previous literature in mind, we attempted to create experiments directly measuring the effects of lecture slide handout annotation vs longhand notetaking. We isolated the encoding function by removing any possibility of revision, thus any differences in memory at testing between those who take notes and those who annotate a slide handout can be attributed to encoding during the lecture. Furthermore, two other between-subjects conditions will be added to this laboratory study; a control condition, included in both experiments, in which participants were not provided with any writing material, thus are left to observe the pre-recorded lecture, and a group who are told to make longhand notes for a friend who is absent from the lecture, a condition that was only included in Experiment 1. The rationale behind this second group is that in order for a friend to understand the notes, they will have to be legible, understandable and written in complete prose, thus will require more cognitive effort and will draw on the generation effect.

Alongside the variable of encoding condition, we added three more within-subjects variables across the two experiments. In Experiment 1, we investigated lecture fluency. This variable was added to see whether either longhand notetaking or lecture slide annotation causes students to be more susceptible to disfluencies during a lecture. Recent findings by Carpenter, Wilford, Kornell and Mullaney (2013) found that, as with desirable difficulty, whilst participants thought that they would remember more for a more fluent lecture, there was no difference in performance when a lecture was delivered disfluently. A possible explanation for this is that the participants must have metacognitively regulated their performance by identifying that the disfluent lecture could be more difficult to follow, thus put in more effort in during the disfluent lecture.

In Experiment 2, we also included the variables of lecture speed and information type. We manipulated lecture speed to investigate whether a faster lecture would be too cognitively demanding for a notetaker, as Kierwa (1985) and Kobayashi (2005) found increased lecture speed to have a negative effect on notetaking. The variable of information type related to the lecture content; whether the information that they were learning was considered to be a fact or a concept. We defined a concept as a process, which includes a series of steps to complete an idea unit, and theorised that the necessity to note down all stages in a concept might be a challenge to the notetaking groups to complete, whilst attending to the lecture, particularly when the lecture was faster paced.

Both the within- and between-subjects variables were tested at two different time intervals; ten minutes after the lecture and one week later to make a distinction between immediate recall (performance) and delayed recall (learning, Sodestrom & Bjork, 2015).

Based on the findings of the literature, we predicted that the desirable difficulties associated with notetaking would mean that participants in those groups would remember more information both immediately following the lecture and after delay, thus would perform better on testing.

Paper 1 [PowerPointLESS? Annotating Lecture Slide Handouts is a Less Effective Encoding Strategy than Longhand Notetaking]

Manuscript Submitted to JEP Applied: 01/07/18

Abstract

Lecture slide handout annotation has largely replaced the once prevalent practice of longhand notetaking. What impact has this had on student learning? In Experiment 1, students viewed two lectures, one presented fluently and the other disfluently, whilst either observing without making any form of notes, annotating handouts, taking notes for themselves or taking notes for a friend before immediate and delayed testing. Students in both notetaking groups out-performed observers and annotators, but there was no difference in performance between the latter groups. This learning benefit from notetaking was not reflected in confidence, suggesting that students are metacognitively unaware of it. In Experiment 2, students viewed two lectures, each consisting of material pertaining to both concepts and facts. One lecture was presented at a regular pace and the other at a faster pace. As with Experiment 1, students who made longhand notes performed better across all of the above conditions than observers and annotators, who did not differ from each other in terms of test scores. These findings suggest that notetaking is more beneficial for memory than lecture slide handout annotation across a wide range of lecture scenarios over both short and long-term periods.

Keywords: Learning, lectures, encoding function, notetaking, PowerPoint

PowerPointLESS? Annotating Lecture Slide Handouts Is A Less Effective Encoding Strategy Than Longhand Notetaking

In the past 25 years, technological advances have changed the way that lectures are delivered in universities (Gier & Kriener, 2009). For example, software such as Microsoft PowerPoint, which is used to present lecture material in a colorful, neatly-organized manner, is now ubiquitous in university classrooms around the world. Compared to traditional lecture formats, this new lecture-delivery format, which often incorporates multimedia as well as textual content, has been generally praised for being more structured (Amare, 2006) and visually interesting (Harknett & Cobaine, 1997) and for improving student lecture experiences (Holzl, 1997).

In this paper, we consider how these changes to lecture delivery may have some potentially negative side effects to learning. Specifically, many instructors make copies of their PowerPoint lecture slides available to students prior to lectures.[1] If students do indeed choose to print out these slides (or have access to them digitally) and bring them along to lectures, it could impact how students encode lecture material, since these notes are typically annotated (in either short, bullet-point or single word format) compared to the complete notes made when no other material is present. This need to take comprehensive, longhand notes, an activity that has kept students busy during lectures

[1] Whilst this practice may vary internationally, it is commonplace in the UK. We contacted 16 of the universities in the South East of England to ask their policy on providing slide handouts. Of the 12 that responded, four of the universities mandated that lecture slides be made available to students to print prior to lectures and a further six strongly encouraged this practice. Only two (13%) had no position on this issue.
[2] The completeness of lecture slide handouts is likely to vary from instructor to instructor such that some lecture slide handouts include only short bullet points whereas others are more complete. However, instructors are typically aware that students use lecture slide handouts for exam review and may ensure that lecture slide handouts contain all the key points from the associated lecture(s).

for centuries, is largely precluded by the presence of these printed lecture-slide handouts, which encourage annotation of the existing printed material which differs in content to longhand notes. For example, in the experiments we report later, providing lecture slide handouts greatly reduced students' notetaking compared to no lecture slide handouts, a finding that is consistent with the previous literature (e.g., Marsh & Sink, 2010).

Di Vesta and Gray (1972) distinguished between the *encoding function* and the *storage function* of notetaking and the same distinction can be applied to lecture slide handouts. The encoding function refers to how the activity of notetaking or annotating handouts affects the encoding of lecture material whereas the storage function refers to how notes or handouts might be used as an external store of information. In terms of storage, lecture slide handouts have some potential advantages over students' handwritten notes (Frey & Birnbaum, 2002). For example, the slides are potentially more accurate, more legible, and more complete than students' notes.[2] Also, lecture slide handouts can contain complex diagrams, figures, photographs, and tables that would be impossible to reproduce in longhand under the time pressure of a lecture (Barbetta & Skaruppa, 1995). These potential storage advantages are clearly important when lecture slide handouts are used as a post-lecture restudy aid, for example, when preparing for examinations.

However, the more interesting question to us is what effect does the presence of lecture slide handouts (and a concomitant reduction in notetaking) have on initial encoding of learned material? This question is interesting because, on the one hand,

---

[2] The completeness of lecture slide handouts is likely to vary from instructor to instructor such that some lecture slide handouts include only short bullet points whereas others are more complete. However, instructors are typically aware that students use lecture slide handouts for exam review and may ensure that lecture slide handouts contain all the key points from the associated lecture(s).

research pre-dating lecture slide handouts (e.g., Cook & Mayer, 1983; Einstein, Morris & Smith, 1985; Peper & Mayer, 1978) suggested that notetaking promotes encoding of lecture information relative to passive observation. Kobayashi (2005) conducted a meta-analysis on the encoding function of notetaking and confirmed this conclusion, particularly for tests involving recall. Thus, from this viewpoint, the presence of lecture slide handouts is deleterious because of it reduces notetaking. On the other hand, these detriments to learning may be offset or even reversed because lecture slide handouts provide an organizational structure that may promote more efficient encoding of lecture material. They may also relieve students from needing to divide their attention between listening and notetaking, which could impair encoding, particularly if lectures are delivered at a fast pace.

The research that is most relevant to the question of whether lecture slide handouts promote or impair learning is that which includes both notetaking and handout conditions Since lecture slide handouts have become more prevalent, several studies have investigated the effect of notetaking versus lecture slide handouts in university classrooms using real academic courses and test scores. Murphy and Cross (2002) and Weatherly, Grabe and Arthur (2003) found that undergraduates who made longhand notes tended to outperform those who relied on lecture slide handouts. Conversely, Babb and Ross (2009) found that students' test scores did not vary regardless of whether lecture slide handouts were provided before or after the lecture. Although the number of notes was not measured by Babb and Ross, if notetaking was more prevalent if lecture slide handouts were provided after the lecture, this null finding contradicts the earlier ones. However, the naturalistic setting of all of these studies makes drawing firm conclusions difficult. Importantly, given that the lecture was part of a university course, students could choose whether to print the handout, even in the condition when it was available. Therefore, compliance with the condition cannot be ensured. Equally, we are unaware of the impact of restudy, given that it is likely that students reviewed the material prior to testing. Finally, although Babb and Ross used two different courses and a counterbalanced design across two semesters in their study, not all aspects of course delivery in each semester were the same. For example, the timing of lecture slide handout provision and the presence of a weeklong administrative break varied. These confounding variables make it difficult or even impossible to properly assess the encoding efficacy of either notetaking or lecture slide handouts and to separate any encoding effects from the effects of storage. We tackle this issue in the current research by not allowing students to revise prior to testing so that we could examine encoding effects free from confounding variables.[3]

Other studies have investigated the value of lecture slide handouts and notetaking in more controlled environments. For example, Marsh and Sink (2010) conducted two laboratory experiments comparing test performance between students given lecture slide handouts and others given none. In their first experiment, participants viewed a pre-recorded lecture in either a handout-present (handout *and* blank paper) or handout-absent condition (no handout, blank paper only) and were tested either immediately or following delay. Marsh and Sink found that for both immediate and delayed testing, participants performed equally across both handout conditions. They concluded that the

[3] In typical academic environments, students will usually have access to both their notes and lecture slide handouts when preparing for exams (i.e., both serve a storage function as well as an encoding function). Hence, focusing purely on the encoding function may seem contrived. However, although examining the encoding function on its own may differ from the typical practice seen in education, it is important to investigate these two functions of notetaking and handouts separately in well-controlled designs so that their separate contributions to exam performance can be ascertained. Our focus here is on the encoding function, but readers may be interested in our other work examining the storage function (e.g., Coria & Higham, 2018b).

presence of a handout allows for more efficient encoding, given that less effort was presumably required to annotate a handout than to make comprehensive longhand notes.. However, this conclusion is compromised in that participants were provided with blank paper to take longhand notes in both groups. Moreover, participants in the handout-absent condition wrote twice as many notes as participants in the handout-present condition. Thus, the handout-present and handout-absent condition differed on more than one factor, rendering the design confounded. In our view, it is not possible to assess the relative benefits of notetaking versus lecture slide handout annotation without the inclusion of a control condition with no handouts or longhand notetaking. We include such a group in both experiments reported here.

Student Perceptions of Lecture Slide Handouts

Before reporting our main studies, we first describe the results of a survey we conducted to determine students' beliefs and preferences about lecture slide handouts. We asked 247 students at our University for their opinions on lecture slide handouts and their usefulness. Eighty-seven percent of those surveyed agreed that they would like an lecture slide handouts to be made available before every lecture, and 59% demonstrated a strong preference for lecture slide handouts by indicating that they printed the slides themselves if the lecturer made them available in advance of lectures. Furthermore, 48% claimed that annotating the lecture slide handout was a sufficient aid to allow them to remember the content of the lecture on later exams. Finally, in accordance with previous research (e.g., Douglas, Douglas & Barnes, 2006), 61% said that they found lectures more difficult when they did not have a lecture slide handout during the lecture.

These finding suggest that students' strong preference for lecture slide handouts may be attributable mostly to them easing lecture comprehension. The fact that almost half of student believed that lecture slide handouts were sufficient to promote long-term retention of lecture content is also telling. A common lay-belief amongst students (and some lecturers) is that making conditions easier for learning encourages enduring memory of what is learned (e.g., Yan, Bjork, & Bjork, 2016). However, a growing body of research applying principles from cognitive psychology to education has found the reverse to be true; generally speaking, it is *difficult* and *slow* learning that promotes long-term memory retention, not easy or fluent learning, a point to which we now turn.

Notetaking as a Desirable Difficulty

Bjork's (1994) concept of *desirable difficulty* suggests that, compared to easy or fluent learning, overcoming difficulties that slow the rate of learning will lead to better long-term memory retention. Bjork and Bjork (2011) argued that *self-generation* is one important desirable difficulty, and there is no paucity of evidence indicating that self-generating information leads to better long-term memory than reading it (e.g., *generation effect,* Slamecka & Graf, 1978; *testing effect,* Roediger & Karpicke, 2006).

Notetaking is, of course, an inherently generative activity in that students must generate their notes from working (or sometimes long-term) memory. Even a disengaged student, who does nothing other than copy slide information verbatim into their notes may benefit from notetaking. Compared to reading, simply *producing* information has been shown to benefit memory (*production effect*, MacLeod, Gopie, Hourihan, Neary & Ozubko, 2010; see also MacLeod & Bodner, 2017). Most research into this effect has focused on vocal production of information, but research by Mama and Icht (2016) found an even greater advantage of writing information compared to vocalizing it.

Conversely, students given lecture slide handouts are more likely to engage in more passive encoding strategies such as observing the lecture, reading the lecture slide handout, or highlighting/annotating information on the lecture slide handout (e.g., Marsh & Sink, 2010). These strategies are generally regarded as having low efficacy (Dunlosky, Rawson, Marsh, Nathan & Willingham, 2013) and require relatively low levels of cognitive effort. By this analysis, then, in term of encoding, we would expect better long-term memory for note takers compared to lecture slide handout annotators.

Experimental Overview

Primarily, our study was designed to investigate how longhand notetaking and lecture slide handout annotation during lecture encoding fare relative to passive observation on immediate and delayed tests of memory, and how any potential differences between these encoding conditions related to students' preferences. Furthermore, to test the robustness of our findings, we manipulated various aspects of the lecture (e.g., the fluency of the lecture delivery in Experiment 1; the pace and type of learning material in Experiment 2) and the test materials (multiple-choice questions in Experiment 1; cued-recall questions in Experiment 2). In Experiment 1, we also manipulated the intended recipient of the notes: participants either made notes for themselves, as is typically the case, or for an absent friend. We hypothesized that students would work harder and take more comprehensive notes in the latter case, which may benefit memory. In support of this hypothesis, Nestojko, Bui, Kornell and Bjork (2014) showed that participants who expected to later teach the material that they were learning engaged in more effective encoding strategies, which benefited their later memory performance. We reasoned that taking notes for an absent friend who would need to read and understand those notes without prior exposure to lecture material (i.e., taking notes to "teach"), could also invoke effective encoding processes. If so, later memory performance may be enhanced relative to taking notes for personal use.

Experiment 1

Experiment 1 was our initial investigation into the relative benefits of notetaking and lecture slide handout annotation on memory for lecture content. The efficacy of notetaking for long-term memory retention is likely to be moderated by a number of variables. One such variable that we examined in Experiment 1 was lecture fluency. Carpenter, Wilford, Kornell and Mullaney's (2013; see also Carpenter, Mickes, Rahman & Fernandez, 2016) found that, although participants believed they would later remember more from a fluent lecture than a disfluent one, there was actually no difference in test performance. This metacognitive illusion was likely caused by the same factor that leads to students' preference for lecture slide handouts: fluent lectures, like lecture slide handouts, make initial encoding of the to-be-remembered lecture material seem easy and, therefore, more likely to be retained. Consequently, compared to disfluent lectures, fluent lectures may cause students to place less importance on notetaking because the lecture material seems self-evident or to "just make sense." In turn, this reduction in notetaking could potentially harm later memory performance in the fluent lecture condition.

Also, lecture fluency may produce a confidence-accuracy dissociation analogous to that observed by Carpenter et al. (2013). However, their dissociation was found with prospective judgments-of-learning (JOLs), which are prone to bias from irrelevant cues (e.g., Hertzog, Hines, & Touron, 2013). In contrast, in Experiment 1, we investigated whether this confidence-accuracy dissociation would generalize to retrospective confidence judgments made at the time of testing. Previous research has suggested that metacognitive ratings made after retrieval are sometimes more accurate than those made prior to retrieval. For example, Busey, Tunnicliff, Loftus, and Loftus (2000) found that JOLs were influenced by factors not related to memory accuracy. Retrospective confidence ratings, on the other hand, were mostly based on the same information as memory accuracy, except in cases where a variable is manipulated during testing. Similarly, JOLs that are delayed for some time after the to-be-judged stimulus is presented tend to be more accurate than immediate JOLs, the so-called *delayed-JOL effect* (Nelson & Dunlosky, 1991). One explanation for this effect is that delaying the JOL encourages people to base the judgment on a covert retrieval attempt rather than some other, less diagnostic cue of later recall performance (e.g., Nelson, Dunlosky, & Narens, 2004).

Furthermore, in addition to a notetaking group who took notes for themselves (as they would in a regular lecture), we also included a separate group who were instructed to take notes as if they were going to give them later to an absent friend who could not attend the lecture. We included this group because we hypothesized that the encoding processes involved in notetaking that may benefit memory would be particularly pronounced. For example, as described earlier, we hypothesized that the requirement to provide information to others would improve students' encoding strategies in the same way that expecting to teach does (e.g., Nestojko et al., 2014). For example, expecting that the notes will be read by others may improve both the organization and completeness of the notes, with less information omitted because it had personal relevance or was deemed easily remembered.

In summary, there were four groups within this experiment: (1) the control group, who passively observed the lecture, (2) the annotation group, who were given an lecture slide handout to annotate during the lecture, (3) the notetaking-for-self (NS) group, who took regular longhand notes on plain paper during the lecture, and (4) the notetaking-for-friend (NF) group who took longhand notes on plain paper for a friend as discussed earlier. Participants were assigned to the encoding groups randomly and viewed two pre-recorded lectures, one presented fluently, and the other disfluently (more detail later). We included both immediate and delayed tests of memory because sometimes desirably difficult tasks do not reveal their benefits until after a delay (Soderstrom & Bjork, 2015). Because we wanted to isolate the encoding function, no students were permitted to review any of the lecture materials prior to the tests.

Method

Participants. An a priori power analysis using G*Power (Faul, Erdfelder & Lang, 2009) was conducted for both the within- and between-subjects effects. The analysis was conducted with power (1 - β) set at .95 to determine how many participants would be needed (per group) to detect an effect size of η_p^2 = .20. This effect size was selected based on the findings of Coria and Higham (2018a), whose between-subjects variable was also encoding method, and Carpenter et al. (2013), whose within-subject variable was also lecture fluency. All of the above found an effect sizes that exceeded η_p^2 = .20. For the between-subjects measures, we assumed that we would be making pairwise comparisons across these groups. Based on this assumption, G*Power indicated that a sample of 21 participants per group was the minimum requisite. For the within-subjects measures, a minimum sample of 12 participants per group was needed

In total, 118 university undergraduate students took part in this research in exchange for financial compensation (£15 for the completion of both sessions) or course credit. The 118 participants were randomly assigned to one of the four encoding groups with the constraint that all participants tested at one time belonged to the same experimental group: control (*n* = 34), annotation (*n* = 29), NS (*n* = 28), and NF (*n* = 27). All 118 participants contributed data to the first session, but 11 participants did not attend the return session one week later (seven from the control group, one from the NS group and four from the NF group). Moreover, two further participants in the NF group provided no confidence ratings after the delay. Thus, the final total sample sizes for analyses involving delayed accuracy and delayed confidence were 107 and 105, respectively. Each participant was verbally asked if they had studied, or been affected by, either topic covered in the lectures and all confirmed that they had not.

Design and materials. The design was a 4 (encoding group: controls, annotation group, NS, and NF) x 2 (lecture fluency: fluent vs. disfluent) x 2 (delay: immediate vs. delayed test) mixed factorial, with encoding group as the only between-subjects variable. All participants viewed two pre-recorded lectures on different topics, one presented fluently and the other disfluently, and completed both an immediate and a delayed multiple-choice test on each lecture. The groups differed in the activity they were required to perform as they watched each lecture: the control group passively observed the lectures; the annotation group received a lecture slide handout to annotate; the NS group and the NF group were both given blank paper and respectively instructed to make notes on the lecture content for themselves, or for an absent friend, while viewing the lectures.

The two prerecorded, 10-min lectures were prepared, one on tuberculosis and the other on influenza. The content related to the prevalence and pathology of each disease. The lectures consisted of Apple Keynote slides with an accompanying prerecorded voiceover to mimic a realistic lecture scenario. See supplementary materials for the slides and a verbal transcript.

The lecture slides were presented in color on an individual computer screen with accompanying voiceover. Each lecture was recorded twice, once with a fluent voiceover and once with a disfluent voiceover. The fluent voiceover was delivered calmly yet enthusiastically, with regular pacing and pauses, and without errors. The disfluent voiceover was delivered more erratically with some errors that were corrected, irregular pacing, and no vocal enthusiasm. These techniques were modeled on those used by Carpenter et al. (2013) in their research on lecture fluency. Depending on group assignment, participants either received a lecture slide handout, consisting of 6 x 2 black-and-white slides which were complete copies of all of the slides shown on the computer screen as part of the lecture, to annotate, blank paper on which to make longhand notes (for themselves or a friend), or no notetaking material (control).

After each lecture, but before the immediate tests, participants engaged in a ten-minute filler task. For this task, participants were asked to write down on a provided sheet of paper as many alternative (i.e., unintended) uses for four common household items (e.g., toothpaste) as possible. Because there were two immediate tests and hence two filler tasks, eight unique items were used, four per filler task.

Four multiple-choice tests were created, two for each topic. Each test consisted of 15 questions with five options (one correct and four incorrect). Questions were based on different parts of the lecture such as epidemiology, mechanisms of infection, and symptoms. An example of a question on the influenza lecture was “what is the name of the RNA strand that causes influenza?” with five potential answers provided: “Chrysoviridae”, “Birnaviridae”, “Orthomyxoviridae”, “Hypoviridae” and “Totiviridae”. The correct answer is “Orthomyxoviridae” and the lecture slide corresponding to this question clearly stated both in visual and auditory form that “influenza is caused by the viral particle Orthomyxoviridae.” For all test questions, see supplementary material.

An assortment of test questions was used. To answer some questions correctly required verbatim memory for slide or voiceover content, whereas others involved slight word changes, but the key information to answer all of the questions in the test was included in the slide or voiceover content. All four tests were of comparable difficulty and each participant completed all four tests. Underneath their chosen answer, participants were asked to indicate their confidence, on a scale of 0-100, that their answer was correct.

The order of lecture fluency type (fluent, disfluent), lecture topic (tuberculosis, influenza), and test version (test 1, test 2), were all counterbalanced across participants, as was the assignment of lecture topic to the fluency variable.

To ensure that the presentations and tests were suitable for our participants, we submitted them to a medical doctor who had lecture experience. He judged the presentations to be at the level of a first-year undergraduate medical-school lecture and of comparable difficulty. The tests, too, were deemed to be of equal difficulty that was at an appropriate level.

Procedure. Each participant viewed the two lectures individually. On arrival, they were seated in front of an Apple iMac computer and, prior to beginning the lectures, they were instructed as to how they should be taking notes in line with their allocated group. For controls, they were instructed to just watch the lecture, for the annotation group, they were instructed to annotate the 6 x 2 slide handout provided, in any space available, for the NS group, they were instructed to make notes on the blank paper and for the NF group, they were told to imagine that a friend could not make the lecture and had asked them to take notes on their behalf. Once ready, they were asked to press play to view the first lecture. They then viewed the lecture whilst wearing headphones with adjustable volume that allowed them to listen to the audio content. Once the lecture ended, participants were presented with a filler task (described earlier) for 10 min, and then they were tested on the content of the first lecture with a paper test. For each question on the test, participants were instructed to both identify the correct answer (from five alternatives) and to indicate their confidence on a 100-point scale (where 0 = no confidence and 100% = full confidence in the test answer),. Following completion of the first test, participants were instructed to view the second lecture, which was followed by a second 10 min filler task involving a set of objects that were different from the first filler task. Finally, participants were tested on the content of the second lecture. To encourage optimal performance, there was no time limit for completion of either of the two tests and participants were instructed to call the experimenter when they were satisfied that they had completed the test to the best of their ability. Participants were then given a partial debriefing and requested to return at the same time exactly one week later. There was no mention of a second test to prevent participants researching the lecture topics between the two tests. Thus, there was no restudy phase within this experiment.

After one week, the participants returned to the same laboratory and were given two more tests, one relating to each of the two lectures viewed a week previously. They were able to choose the order in which they completed the tests and, as previously, they were instructed to choose one answer for each question and to indicate their confidence. Again, both tests were self-paced. The immediate and delayed tests were counterbalanced across participants. Finally, before leaving, participants were fully debriefed and thanked for their time. If participants did not attend the second session, they were sent full debriefing information by email.

Ethics approval. The study was reviewed and approved by our Ethics Committee and the Research Governance Office. All participants were consenting adults aged over 18 years old and all gave informed consent for their data to be included in the study and any publications that followed from it. Following completion of the study, participants were fully debriefed.

Results and Discussion

Order effects. To determine whether there was a difference in participants' engagement between the first and second lecture and test, we conducted a 2 (test order: first, second) x 2(test interval: immediate, delayed) between-subjects Analysis of Variance (ANOVA) on test accuracy. Results showed a main effect of delay on test accuracy (discussed below), but no main effect of order, nor an interaction, largest $F(1,106) = 2.82$, $p = .10$, $\eta_p^2 = .026$. These results demonstrate that the order of the material presented did not affect their test scores, thus their engagement or motivation during the second lecture appears to be the same as the first.

Accuracy. Table 1 shows mean test scores (as a percentage) and confidence intervals (CIs) as a function of retention interval, lecture delivery, and encoding group. A 4 (group: control, annotation, NS and NF) x 2 (lecture fluency: fluent, disfluent) x 2 (test interval: immediate, delayed) mixed-model Analysis of Variance (ANOVA) on accuracy, with group as the only between-subjects factor, revealed a main effect of delay, $F(1,103) = 143.78$, $p < .001$, $\eta^2 = .58$, $d = 2.36$. Unsurprisingly, participants scored significantly higher when tested immediately following the lecture (*M* = 65.46, *95% CI* [63.43, 67.49]) compared to when tested one week later (*M* = 52.27, *95% CI* [50.18, 54.37]). More importantly, there was also a main effect of group, $F(3,103) = 9.50$, $p < .001$, $\eta_p^2 = .22$. The overall group means are shown in the rightmost column of Table 1. Results of Fisher's Least Squared Difference (LSD) tests revealed that whilst the NS group and the NF groups' scores did not significantly differ from each other, both groups scored higher than both controls and the lecture slide handout annotation group. However, no significant difference was found between the control group and the annotation group. No other main effect or interaction was significant from the analysis of accuracy, largest $F(3,103) = 1.66$, $p = .18$, $\eta_p^2 = .05$.

Confidence. Table 2 shows the means (and CIs) for the confidence ratings participants assigned to test answers. An analogous 4 x 2 x 2 mixed-model ANOVA on mean confidence for test answers revealed a main effect of delay, $F(1,101) = 219.38$, $p < .001$, $\eta^2 = .69$, $d = 2.95$. Predictably, participants had significantly higher confidence in their test answers on the immediate tests (*M* = 65.46, *95% CI* [63.43, 67.49]) compared to the delayed tests (*M* = 47.67, *95% CI* [44.21, 51.14]). This decrease in confidence for delayed answers corresponded to the decrease in accuracy described earlier. More interestingly, the main effect of lecture fluency was also significant, $F(1,101) = 5.38$, $p = .022$, $\eta^2 = .05$, $d = .46$. Participants were more confident in their test scores when the lecture was delivered fluently than when it was delivered disfluently. There were no other significant effects from the ANOVA on confidence, highest $F < 1$, including no significant main effect of group.

The content of notes and annotations. To discover how notes and annotations were affected by encoding task and lecture fluency, and how they related to immediate and delayed test performance, we analyzed their content. The control group was not included in any of these analyses because they wrote nothing during the lectures.

We first computed a *Word-Count* variable, which was simply the number of words that participants wrote in their notes during the lectures, regardless of content. We investigated this variable because the act of taking copious notes could conceivably focus attention on the lecture content, leading to better learning. Second, we computed a second variable that measured the quality of the notes rather than their quantity. Specifically, we searched participants' notes for information that could be used later to answer test questions correctly. This *Correspondence* variable (so-named because it reflects the extent to which participants' notes correspond with the key lecture points that were tested) was the number of test questions (15 per test) for which the correct answers had been written earlier in the notes. The correct information could have been either copied verbatim or paraphrased from the lecture slides and/or lecture voiceovers. Correspondence was not computed for the annotation group because the handouts already contained all the information necessary to answer the test questions correctly, so whether or not the information was repeated in the annotations was not informative. Both the Word-Count and Correspondence variables were computed separately for fluent and disfluent lectures and for the different encoding group. Retention interval was not relevant in this analysis; the amount written in the notes (Word-Count) was the same regardless of whether the test was immediate or delayed. Also, counterbalancing the immediate and delayed tests across participants meant that the tests were the same at the group level. Hence, the Correspondence scores for immediate and delayed tests were logically equivalent.

The results of the Word-Count analysis are shown in Table 3. The data were analyzed with a 3 (group: annotation, NS, NF) X 2 (lecture fluency: fluent, disfluent) mixed-model ANOVA, which revealed a main effect of group, $F(2,67) = 22.86$, $p < .001$, $\eta_p^2 = .41$. Neither the main effect of lecture fluency nor the interaction was significant, largest $F < 1$. As a follow up to the main effect of group, Fisher's LSD tests revealed that both the NS and NF groups scored significantly higher than the annotation group, but their scores did not significantly differ from each other. Turning now to the Correspondence analysis (Table 4), a 2 (group: NS, NF) x 2 (lecture fluency: fluent, disfluent) mixed-model ANOVA revealed no main effects or interactions, all *F*s < 1. The absence of any effect of lecture fluency in either the analysis of Word Count or Correspondence suggests that, contrary to our hypothesis, fluent versus disfluent lecture style had no effect on either the quantity or quality of our participants' notes.

We were also interested in potential relationships between the quantity and quality of the notes on the one hand and later immediate and delayed test performance on the other. Did writing a lot and/or writing information pertaining to correct test answers during the lectures help later test performance? To find out, we computed correlations (Pearson's *r*s) between Word-Count and Correspondence on the one hand and total immediate and delayed test performance on the other. Because lecture fluency had no effect on test performance, we pooled the fluent and disfluent test scores. Again, the annotation group was not included in this analysis because they wrote much less than in either of the notetaking groups, which could have produced spurious correlations.

The results showed a clear pattern. Correspondence was a significant predictor of both immediate test performance, $r[54] = .38$ $p = .005$, and delayed test performance, $r[30] = .36$, $p = .011$. In contrast, analogous correlations between Word-Count variable on the one hand and immediate and delayed test performance on the other were not significant, largest absolute $r[34] = .09$, $p = .510$. Thus, writing the correct answers to later test questions in the notes was a good predictor of later test performance, regardless of whether the test was taken immediately or after a one-week delay. Conversely, the sheer amount written (Word-Count) did not predict later test performance at all.

In summary, there were a number of important findings obtained in Experiment 1. First, longhand notetaking, whether notes were created for the self or others, conferred both short- and long-term learning advantages compared to passive observation. The sheer amount written in the notes was not a good predictor of later test performance, suggesting that notetaking does not benefit learning simply by maintaining attention during the lecture. Instead, generating key ideas in the notes that were later tested predicted performance both immediately and after a delay. However, in contrast to notetaking, slide annotation conferred no short- or long-term learning advantages at all compared to passive observation.

At a metacognitive level, the learning advantage of notetaking was not reflected in subjective confidence, but confidence *was* sensitive to the fluency with which the lecture was delivered. This pattern is exactly opposite to the effect of these variables on accuracy (i.e., encoding group had an effect whereas lecture fluency did not). Together, these results suggest that confidence responded mostly to the ease with which lecture information was encoded rather than to desirable difficulties of the encoding task.

Experiment 2

In Experiment 1, we found that both notetaking groups outperformed both the lecture slide handout annotation group and the passive observation group. In Experiment

2, we sought to replicate this important finding and test its robustness in a variety of ways. First, we investigated whether the notetaking advantage generalized to different lecture materials. Whereas in Experiment 1 we examined memory for lectures on medical topics (tuberculosis and influenza), in Experiment 2, we investigated memory for lectures on natural science (depletion of the ozone layer and acid rain). These topics, like the medical ones used in Experiment 1, were suitable for creating lectures that were rich in both fact- and concept-based information (see later).

Second, we retained the two test intervals used in Experiment 1 (i.e., participants were tested both immediately following the lecture and after a week-long delay) but we used a cued-recall test format as opposed to the multiple-choice format used in Experiment 1. Multiple-choice questions can be answered correctly in a variety of ways that which arguably do not represent legitimate learning. For example, they can be selected strategically (e.g., by eliminating implausible alternatives), or, correct answers can be selected on the basis of vague feelings of recognition, compared to recall which is considered to require fewer cues to lead to retrieval, thus memory (Gillund & Shiffrin, 1984). In some cases, correct answers can be selected through pure guesswork. In contrast, these paths to good performance are much less likely to be present with cued-recall tests. Instead, explicit retrieval of lecture material is the primary basis of good performance, which unquestionably represents legitimate learning (Carpenter, Pashler & Vue, 2006). We predicted that we would continue to observe the benefits of notetaking in Experiment 2 despite the change in test format because generative activities, such as those involved in longhand notetaking, have been shown in other research to facilitate explicit retrieval (e.g., generation effect: Slamecka & Graf, 1978; testing effect: Roediger

& Karpicke, 2006). In this vein, Kobayashi (2005) found that the effect of notetaking was larger for recall tests than for recognition (multiple-choice) tests.

After findings in Experiment 1 showed that the fluency of the presentation of a lecture did not mitigate the benefit of longhand notetaking, we decided to test other aspects of lecture content and delivery would limit the benefits of notetaking seen in our first experiment. Therefore, we replaced the lecture fluency variable manipulated in Experiment 1 with a *lecture delivery speed* variable in Experiment 2: participants viewed one lecture at a regular pace and another at a faster pace. We included this variable with the intuition that it would moderate the benefits of notetaking; whilst a faster paced lecture could prove less of an obstacle for a student annotating a slide handout. Any student who has tried to take notes in situations where the lecturer is speaking too quickly knows that notes on some portions of the lecture can be missed while trying to keep up. Despite this, some research (see Barabasz, 1968; Aiken, Thomas & Shennum, 1975) found that increasing the speed of a lecture was not accompanied by a reduction in test scores, there was no mention in this research of the kind of strategies students adopted during encoding to cope with the fast pace. Indeed, lecture pace and its effect on notetaking has largely escaped attention in the literature. Given that notetaking is time consuming compared to lecture slide handout annotation or passive observation, attending to the material and making notes in a faster-paced lecture could strain the demands of working memory (Piolat, Olive & Kellogg, 2005). If so, the desirable difficulties associated with notetaking could become insurmountable obstacles, limiting its efficacy. Indeed, Kobayashi (2005) and Kierwa (1985), suggested that notetaking may be affected by presentation speed. Also, Van Meter, Yokoi and Pressley (1994) found that students complained that lectures presented too rapidly did not allow them to make generative notes. However, these findings were not supported by evidence that test performance was impaired, so there is a danger that the basis of these complaints was a false heuristic that easy, fluent learning improves later memory performance (e.g., Hertzog, Dunlosky, Robinson, & Kidder, 2003).

Another new variable that we included in this experiment to test the robustness of notetaking benefits was *information type*. Specifically, when creating both lectures, we separated the content into *concepts* and *facts*, which has been shown in previous research on notetaking to be a potentially important distinction (e.g., see Mueller & Oppenheimer, 2014). The literature varies on what constitutes a definition of a “concept” vs a “fact” in learning (for example, Gagne, 1972, defined concept learning as a students’ ability to figure processes in their own words, and provide their own examples), and Shapiro et al. (2017) concluded that it is not possible to provide a single definition distinction as to what constitutes a concept in learning. For the purposes of our research, we define a concepts as material which involved a series of steps to complete an idea unit (e.g., the formation of acid rain), whereas facts related to single idea units that could be understood on their own (e.g., the names of the household appliances which contribute to the depletion of the ozone layer). We selected these definitions as we were interested in determining whether the efficacy of longhand notetaking as a strategy would be limited by the type of information covered in a lecture. A concepts composed of a series of steps which must be followed to be able to understand the idea as a whole, could be more difficult for participants engaged in notetaking, particularly in the faster-paced lecture, thus they may not be able to attend to the lecture whilst taking notes on all of the steps. Additionally, notetaking participants may not be able to progress beyond a certain step if an earlier step had been missed. These problems could cause participants to potentially fail at encoding some of the necessary information due to a progressive cognitive overload. This failure could affect their later test performance, compared to participants provided with the structured lecture slide handout to annotate, which would allow them to “get back on track” if any part in a concept was missed.

To examine these issues, we compared three groups in Experiment 2: a control group, who passively observed the lecture, an annotation group who were given an lecture slide handout to annotate, and a NT group, who took regular longhand notes on plain paper. Because the NT and NF groups did not differ in Experiment 1, we excluded the NF group in Experiment 2. As in Experiment 1, participants in the various encoding groups viewed two pre-recorded lectures, one presented at a regular pace and one presented at a faster pace (25% increased speed). Both lectures consisted of an equal mix of concept- and fact-based idea units. At testing, participants answered 12 cued-recall questions per lecture and completed both an immediate and a delayed test (1 week). Because our main focus was to test the robustness of our findings with test performance in Experiment 1, we dropped the confidence variable in Experiment 2. Finally, as in Experiment 1, no review of the lecture materials was permitted in order to isolate the encoding function of participants’ allocated strategy free from confounding variables.

Method

Participants. A power analysis using G*Power (Faul, Erdfelder & Lang, 2009) was conducted for both the within and between subjects effects using the same criteria as Experiment 1, thus the same minimum sample sizes (largest prerequisite group size as shown in Experiment 1 for between-subjects measures = 21 participants per group) were adhered to.

In total, 82 undergraduate students took part in this research in exchange for course credit. The 82 participants were randomly assigned to one of three encoding groups with the constraint that all participants tested at one time belonged to the same experimental group: control (*n* = 27), annotation (*n* = 26) or NT (*n* = 29). Seven participants were excluded (two from the control group, four from the annotation group and one from the NT group) after indicating that they had prior knowledge on one or both of the lecture topics, a pre-determined exclusion criterion, and a further two participants (one from the annotation group and one from the NT group) were excluded as their scores deviated by more than two standard deviations from the mean score of their group. The remaining 73 participants contributed data to the first session, but 11 participants did not attend the return session one week later (two from the control group, two from the annotation group and seven from the NT group). Thus, the final total sample size for analyses involving delay was 62.

Design and materials. The design was a 3 (encoding group: control, annotation, NT) x 2 (information type: concept vs. fact) x 2 (lecture pace: regular vs faster) x 2 (test delay: immediate vs. delayed) mixed factorial, with encoding group as the only between-subjects variable. The materials consisted of two pre-recorded lectures on natural science, the first on the depletion of the ozone layer and the second on acid rain. The lectures presented in colour and were recorded using Apple Keynote were and comprised of slides with an accompanying prerecorded voiceover as in Experiment 1. Both lectures consisted of six conceptual idea units and six factual idea units. An example of a conceptual idea unit was a description of the process through which acid rain causes damage to marble statues via two chemical reactions. An example of a factual component was a list of the countries most affected by depletion in the ozone layer. For the slides and verbal transcript of both lectures in Experiment 2, see supplementary materials.

To implement the lecture-pace factor, each lecture was recorded twice, once with a regular pace and once at a faster pace. Both the regular- and faster-paced lectures on each topic consisted of the same material (both slides and voice over), but the faster-paced lecture was 8 min in length (with the pace of the ozone and acid rain lectures at an average of 87- and 84-wpm respectively) whereas the regular-paced ozone and acid rain lectures were 10 min (an average of 70- and 67-wpm respectively). The order of lecture pace (regular, faster), and lecture topic (ozone depletion, acid rain), was counterbalanced across participants, as was the assignment of lecture topic to the lecture-pace variable.

As in Experiment 1, the groups differed in the activity they were required to perform as they watched each lecture: the annotation group received a lecture slide handout to annotate; the NT group was given blank paper and instructed to make notes as they would during a regular lecture; and the control group passively observed the lectures and were provided with no materials.

After each lecture, but before the immediate tests, participants engaged in a ten-minute filler task. This task was the same as in Experiment 1.

To assess performance, four cued-recall tests were created, two for each lecture. Each test consisted of 12 questions. Six questions relating to conceptual aspects of the lecture and six questions relating to factual aspects. Underneath each question, a single blank line was provided on which participants could write their answer. The cued-recall questions varied in required answer length, with some questions requiring a single word, number, or formula for an answer, and others requiring a maximum of a sentence for an answer. As in Experiment 1, an assortment of test questions was used, although all of the information required to answer each question was presented in both the voiceover and lecture slides. An example of concept question from the acid rain lecture was “When Nitrogen Dioxide reacts with the hydroxyl radical, what acid is formed?”. The correct answer is “Nitric Acid” and the lecture slide corresponding to this question clearly states this information both in the slide and verbal transcript. This question is part of the wider concept of how acid rain is formed, explained in steps over a series of slides (see supplementary material). An example of a fact question from the ozone lecture was “Other than aerosols, name a household item that contains Cluoroflurocarbons?” and there were several potential correct answers, as outlined in the lecture, such as “aerosols” or “refrigerators”, any of which would have been accepted as correct.

All four tests were of comparable difficulty and all participants who completed the entire experiment completed all four tests, which were counterbalanced for immediate and delayed testing. The assignment of the two tests per lecture topic to the immediate vs. delayed test conditions was counterbalanced across participants.

Procedure. Each participant viewed the two lectures individually. On arrival, participants were given a screening questionnaire in which they ticked whether they had previously studied or been affected by either of the lecture topics. They were then seated in front of an Apple iMac computer and, as with Experiment 1, briefed as to what to do for their encoding group. Once ready, they were asked to press play to start the first lecture, which they viewed whilst wearing headphones with adjustable volume to allow them to adequately hear the audio content. Once the lecture ended, participants were presented with a filler task (described above) for 10 min, and then they were tested on the content of the first lecture with a paper test. For each question on the test, participants were instructed to write the correct answer, and were informed that their written answer should not exceed the amount of space provided for each question (two lines on a vertical A4 page, see supplementary materials for test questionnaires). For each test, participants were encouraged to guess (i.e., not leave blanks) if they were unsure of any of the answers.

Following completion of the first test, participants were instructed to view the second lecture, which was followed by a second ten-minute filler task involving a different set of objects. Finally, participants were tested on the content of the second lecture. To encourage optimal performance, there was no time limit for completion of either of the two tests and participants were instructed to call the experimenter when they were satisfied that they had completed the test to the best of their ability. Participants were then given a partial debriefing and requested to return at the same time exactly one week later. Their notes were collected prior to their leaving the first session, thus they were not able to restudy them prior to the second session. As in Experiment 1, there was no mention of a second test to prevent participants researching the lecture topics between the two tests. Thus, there was no review phase. Tests scores were marked with participants scoring 1 point for each correct answer and 0 points for each incorrect answer. There were no part marks awarded. Since no questions needed more than one idea unit in an answer, any response provided that did not fully answer the question was marked as incorrect.

After one week, the participants returned to the same laboratory and were given two more tests, one relating to each of the two lectures viewed a week previously, which were different from those written a week earlier. They were able to choose the order in which they completed the tests. Again, both tests were self-paced. Finally, before leaving, participants were fully debriefed and thanked for their time. If participants did not attend the second session, they were sent full debriefing information by email.

Ethics approval. The study was reviewed and approved by the University of Southampton Ethics Committee and the Research Governance Office. All participants were consenting adults aged over 18 years old and all gave informed consent for their data to be included in the study and any publications that followed from it. Following completion of the study, participants were fully debriefed.

Results and Discussion

Order effects. To ensure that there were no order effects, a 2 (test order: first, second) x 2 (test interval: immediate, delayed) between-subjects Analysis of Variance (ANOVA) on test accuracy was conducted. Results showed a main effect of delay on test accuracy (discussed below), but no main effect of order, nor an interaction, largest $F(1,61) = 3.13$, $p = .08$, $\eta_p^2 = .05$. These results are in line with our findings in Experiment 1 and demonstrate that the order of the material presented did not affect their test scores.

Accuracy. Tables 5 (immediate test) and 6 (delayed test) show mean percentage test scores as a function of lecture pace, information type, and encoding group. A 3 (group: control, annotation, NT) x 2 (lecture pace: regular, faster) x 2 (information type: concept, fact) x 2 (test interval: immediate, delayed) mixed-model ANOVA on the test scores (with group as the only between-subjects factor) revealed a main effect of test interval, $F(1,59) = 159.51$, $p < .001$, $\eta^2 = .73$, $d = 3.29$. As expected, participants scored significantly higher when tested immediately following the lecture ($M = 52.79$, *95% CI* [49.25, 56.34]) compared to when tested one week later ($M = 34.16$, *95% CI* [31.16, 37.16]). There was also a main effect of information type, $F(1,59) = 74.13$, $p < .001$, η^2=

.56, *d* = 2.24. Participants scored significantly higher when tested on material pertaining to facts presented in the lecture (*M* = 52.18, *95% CI* [49.15, 55.21]) compared to concepts presented in the lecture (*M* = 34.77, *95% CI* [30.75, 38.80]).

More importantly, as shown in the far-right columns of Tables 5 and 6, there was a main effect of group, $F(2,59) = 5.62$, $p = .006$, $\eta_p^2 = .16$. Results of Fishers LSD tests revealed that whilst the control and annotation groups' scores did not significantly differ from each other, the NT group scored significantly higher than both the control and annotation groups.

The ANOVA also revealed an interaction between test interval and information type, $F(1,59) = 8.42$, $p = .005$, $\eta^2 = .13$, $d = .76$. Participants showed more forgetting between immediate and delayed testing for information pertaining to facts (*M* = 63.78, *95% CI* [59.86, 67.69] and *M* = 40.59, *95% CI* [36.90, 44.28], respectively), than for information pertaining to concepts, (*M* =41.81, *95% CI* [36.88, 46.74] and *M* = 27.74, *95% CI* [23.73, 31.75], respectively).

A further interaction between lecture pace and material was found to be significant, $F(1,59) = 4.32$, $p = .042$, $\eta^2 = .07$, $d = .54$. Participants showed virtually no difference between faster- and regular-pace lectures for performance on questions relating to facts, (*M* =52.36, *95% CI* [48.63, 56.10], and *M* = 52.00, *95% CI* [48.44, 55.56], respectively), but showed reduced performance for faster lectures compared to regular lectures on questions related to concepts (*M* =31.88, *95% CI* [26.75, 37.01] and *M* = 37.67, *95% CI* [32.97, 42.36], respectively). No other main effect or interaction was significant from the analysis of accuracy, largest $F(2,59) = 2.47$, $p = .093$, $\eta_p^2 = .08$.

The content of notes and annotations. As with Experiment 1, notes and annotations from the annotation and notetaking groups were analyzed to attempt to identify how they related to test performance. Both the *Word-Count* and *Correspondence* variables were computed in the same way as in Experiment 1.

The results of the Word-Count analysis were analyzed with one-way ANOVA. This revealed a main effect of group, $F(1,45) = 58.80$, $p < .001$, $\eta^2 = .57$, $d = 2.29$. Participants in the annotation group wrote significantly fewer notes than the notetaking group (M = 41.60, *95% CI* [41.61, 83.35] and M = 169.31, *95% CI* [150.55, 180.06], respectively). However, unlike Experiment 1, the computed correlations (Pearson's *r*s) between word count and correspondence compared to immediate and delayed test performance for the notetaking group showed that neither word count nor correspondence were significant predictors for either immediate or delayed performance, highest $r[19] = .28$, $p = .26$. Whilst we also found in Experiment 1 that word count was not a predictor of immediate or delayed test score, we expected that the correspondence variable would have acted as a predictor for test performance. However, this result is likely due to the small sample size available when only analyzing the notetaking participants who attended both immediate and delayed parts of the experiment.

The results of Experiment 2 indicate that the learning benefits of longhand notetaking are highly robust. As in Experiment 1, the NT group out-performed both the annotation group and passive observation group. This result was replicated despite changes to lecture content (natural science topics instead of medical topics) and test format (cued-recall testing instead of multiple-choice testing). Furthermore, the group factor did not interact with any of the other factors that we manipulated in this experiment (test interval, lecture pace, or information type). Indeed, Tables 5 and 6 indicate that for both immediate and delayed testing, the performance advantage for the NT group over the other groups held even for concepts in the faster-paced lecture, which would have been the most demanding situation in which to take complete longhand notes. It is also noteworthy that, just as in Experiment 1, lecture slide handout annotation conferred no learning benefits at all over passive observation. In fact, mean test scores in the annotation group were descriptively *lower* than in the passive observation group at both immediate (Table 5) and delayed (Table 6) testing.

General Discussion

The primary finding from both Experiments 1 and 2 was that those students who engaged in longhand notetaking while viewing the lectures scored significantly better on tests of the lecture material compared to students who passively observed the lectures (control groups) or students who were provided with lecture slide handouts to annotate (annotation groups). Furthermore, the notetaking advantage held across a wide range of potential factors that potentially could have reduced it. We observed the advantage regardless of whether (a) the test interval was long or short (Experiments 1 & 2), (b) students took notes for themselves or an absent friend (Experiment 1), (c) the lecture was presented fluently or disfluently (Experiment 1), (d) the lecture pace was faster or regular (Experiment 2), (e) concepts or facts were tested (Experiment 2), (f) the test format was multiple-choice (Experiment 1) or cued recall (Experiment 2) (g) the lecture material was on medicine (Experiment 1) or natural science (Experiment 2). Taken together, these results suggest that the notetaking learning benefit obtained in our research is a highly robust phenomenon.

Another noteworthy finding from our research is that annotating lecture slide handouts conferred no learning benefits at all compared to passive observation. In fact, in

Experiment 2, test scores were descriptively *lower* in the group that annotated lecture slide handouts than in the group who did nothing but sit back and listen. This finding stands in direct contrast to students' belief, outlined in the Introduction, that lecture slide handouts are beneficial to learning lecture material. Thus, it appears that we have uncovered yet another example of the false heuristic "easy learning leads to better memory" that is pervasive in the literature (e.g., Benjamin, Bjork, & Schwartz, 1998; Carpenter et al., 2013; Hertzog et al., 2003; Kornell & Bjork, 2008; Rhodes & Castel, 2008; Yan et al., 2016).

Our experiment reveals results that appear to deviate from analogous comparisons in Marsh and Sink (2010, Experiment 1). Whereas we observed better performance in the notetaking groups in both of our experiments, they found no advantage of notetaking. Marsh and Sink's study is one of a few that has both isolated the encoding function in a controlled design and tested both a lecture slide handout annotation group and a longhand notetaking group. However, in our view, these discrepant results may not be as critical as they first appear. As we noted earlier, in their "handout-present" condition, participants were also provided with blank paper, with which to make longhand notes. An analysis into note content within this experiment revealed that participants did use the blank paper, although wrote half as many notes on average (p.697). This leaves the effect of the handout inconclusive, as there is no way of deciphering which method (using the handout or making longhand notes) was actually used by the participants in the "handout present" condition. Indeed, it is entirely possible that participants in this condition paid little attention to the handout, but acknowledged its awareness, thus made fewer notes, but still noted the key material down, so were able to perform equally with the "handout-absent" group.

Despite this, our study is the first to demonstrate such low efficacy of lecture slide handouts. We believe that a contributing factor to that low efficacy is that, unlike most previous studies, we isolated *encoding* processes; students were not permitted to use either the lecture slide handouts or their notes to study the material prior to the tests. In our view, there is a need for more studies with tight experimental control. As discussed earlier, much of the literature on notetaking is difficult to interpret because of confounding variables and/or failure to implement the necessary controls. Only by isolating the encoding function and comparing performance to the appropriate baseline is it possible to ascertain what effect notetaking has on encoding. Previous experiments, such as Babb and Ross (2009), who used real lectures to investigate the benefit of notetaking vs slide annotation, were not able to control for these factors, and found null effects, likely due to the various uncontrolled external factors, one of which being amount of restudy between courses.

Had restudy been permitted, it is possible that the test performance of the annotation groups could have equaled or surpassed that of the notetaking group. Like some educators, we ensured that our lecture slide handouts contained all of the critical information contained in our lectures. Complete lecture slide handouts of this sort likely would have provided a better restudy tool than even the best students' notes, potentially nullifying any encoding benefits caused by notetaking during encoding. Although we do not know of any current literature investigating whether restudy mitigates any previously-seen benefits of longhand notetaking compared with handout annotation, research has indicated that restudy that benefits provided by certain encoding strategies are attenuated by restudy (i.e., Storm, Friedman, Murayama and Bjork, 2014, found that restudy overrides any effects of retrieval practice during the encoding phase of learning).

If the encoding benefits of notetaking are nullified by restudy, why did we choose to isolate the encoding function? We have two responses to these questions. Firstly, it is important not to assume that students will restudy all material. Restudy relies on students to assess what they already know, and what they need to revise, and research such as Rawson and Dunlosky (2007) has found that students tend to over-estimate what they have learned. This is likely to lead to less restudy than is needed, thus the encoding benefit of notetaking would enhance memory for the untested material. Secondly, we believe that restudy does not abolish the encoding benefits of notetaking. Coria and Higham (2018b) permitted students to review material prior to being tested and have implemented test delays of several weeks to more closely simulate typical university assessment. In contrast the current experiments, under these circumstances, we found that the test performance of the notetaking and annotation groups were both better than the passive observation group and differed very little from each other. However, that is not the whole story; although there was no difference in test performance, there was a difference in the amount of time students needed to revise for the exam. Specifically, compared to the notetaking group, the annotation group needed longer to ready themselves for the test, presumably because they needed to relearn more of the material than the notetaking group. Thus, it appears that enhanced lecture encoding caused by notetaking is beneficial under a wide variety of circumstances, but if restudy is allowed, those benefits are not always seen on test performance. Rather, they are seen in more subtle measures of performance, such as the amount of time needed to prepare for tests.

Although memory was improved by notetaking, Experiment 1 showed that students had little insight into this benefit; whereas group exerted a significant effect on test accuracy, confidence ratings were statistically equivalent between the groups. This null result was not attributable to the confidence ratings generally being an insensitive measure. Students' confidence *was* affected by a variable that had no impact on memory accuracy: lecture fluency. This result replicates and extends Carpenter et al.'s (2013) finding that students' metacognitive ratings overestimate the impact that poor lecture delivery has on learning. However, in some sense, the fact that we observed this overestimation with retrospective confidence judgments rather than JOLs as used in Carpenter et al.'s research makes the result even more impressive for two reasons. First, the temporal separation of the lecture and retrospective judgments is greater than with prospective judgments, making the disfluent nature of the lecture less available as a cue at the time the confidence judgment is made. Second, unlike prospective judgments made during learning, which are open to the influence of all kinds of cues, bias from irrelevant cues such as lecture fluency can be tempered by retrieval (memory) strength at the time of testing (e.g., Siedlecka et al., 2016).

The content analysis we conducted on students' notes in Experiment 1 suggested that the sheer number of notes taken was not a critical to good test performance. This finding suggests that the benefits of notetaking are not solely attributable to the maintenance of attention throughout the lecture. Instead, the important factor was whether or not students produced the key information in the notes that was included at testing. The finding that writing correct answers in their notes is important is reminiscent of the testing effect, whereby producing correct answers on practice tests greatly facilitates retrieval of those answers on later tests (see Roediger & Karpicke, 2006 for a review). As noted earlier, taking notes could be considered a form of short-term test in that students must practice retrieving lecture information from working memory, or even long-term memory in some cases. In this vein, Coria and Higham (2018a) found that if notetaking was delayed somewhat by only permitting it during brief lecture pauses during which students listed key points from the preceding lecture segment (i.e., retrieval practice), then later test performance was excellent.

It is important to acknowledge that whilst these results indicate that the encoding method of longhand notetaking is advantageous, there are various differences between the lecture series used in our experiment and typical university lectures. Notably, the length of our lectures which, at between 8-min and 10-min in length, was shorter than would be typically expected in a university lecture, which would usually be between 45-min and 90-min in length. This was a practical decision, based on requisite to recruit many participants, who were required to view two lectures, and who we believed would not attend to lectures of longer than 30-min without the motivation or interest of them being part of their university course. However, we do not believe this affected our results, as many research papers (such as Spzunar, Kahn and Schachter, 2013) have used shortened lectures within their experiments. Furthermore, with the increased use of online academic programmes such as Udemy, ItunesU, academic podcasts and TEDtalks amongst the population in general, particularly students, it appears that lectures are moving towards a more modular approach, with longer lectures being broken down into shorter topics, such as the ones seen in our experiments. Research such as Copley (2007) found that most students respond positively to online podcasts which can be downloaded onto a mobile phone and tend to be shorter in length. Furthermore, the Open University UK is currently puts all of its online content on itunesU, which typically consists of various short lectures, in order to reach more of the population through a single software programme (Law, Perryman & Law, 2013) The virtual delivery of our lecture, also varies from the "live" lecture experience. However, more and more university courses are available online (some exclusively online, see the previous Open University example), and more students are electing this method to view their lectures, where available (Gorissen, Van Bruggen & Jochems, 2012).

Furthermore, the act of longhand notetaking during encoding will not necessarily lead to improved memory amongst all students. Bjork & Bjork (2011) emphasised the importance of *desirable* difficulty, stating that once the process is no longer optimising encoding, and the learner is unable to overcome the obstacles necessary to process the information, the difficulty becomes undesirable. What is a desirable difficulty for some might become an insurmountable obstacle for others. Certain student groups (such as second language students and those with a learning difference such as dyslexia) might benefit from partial structure, such as lecture slide handouts with outlines and illustrations to aid notetaking (Bui & McDaniel, 2015). Indeed, Markovits and Weinstein (2018) discuss how “guided notes” - instructor-prepped notes with either blank spaces for students to insert missing information or preset questions or instructions for student to complete during the lecture (e.g. provide a definition for a term), encourage generative activities during lectures, which promote learning, within the structure of a handout. Markovits and Weinstein also found that these guided notes can help students with organisation during a lecture, and this concept has been well received with students across the world (see Avval, Jarahi, Ghazvinni,& Yousefi, 2013). However, a meta-analytic review investigating guided notes found inconclusive findings for their efficacy in post-secondary settings when compared to regular notetaking (Konrad, Joseph & Everleigh, 2009), Stark-Wroblewski, Kreiner, Clause, Edelbaum and Ziser (2006) found no significant difference in performance whether participants were given regular lecture slide handouts (with complete content, as were used in our experiment), or guided notes.

Our recommendation for instructors based on the results in this paper is to encourage students to take notes during lectures and not allow access to complete lecture slide handouts (i.e., lecture slide printouts containing all the information delivered in the session) during the lecture. Given our results, this advice holds regardless of the quality of the lecture delivery (be it too quickly delivered, or delivered somewhat disfluently), or the type of material covered in the lecture (concepts or facts).

Conclusion

Our study adds to a growing body of research aimed at uncovering ways to enhance student learning and long-term retention in educational settings by applying principles from cognitive and metacognitive psychology. Our research differs somewhat from much of this other work in that we have focused on how best to *encode* lecture material rather than on how best to *revise* it (e.g., Dunlosky et al., 2013). Nonetheless, both lines of research have demonstrated that many of the intuitions that students have about what is good for learning are at odds with the reality. These false intuitions have led to increasing pressure on educators to make their lecture slides available to students well in advance of lectures so that learning is made as easy as possible; indeed, some universities in the UK mandate it. Our research questions whether this is actually best practice. Perhaps, instead, lecturers should persuade students to overcome the desirable difficulty of notetaking.

Paper 1 - Tables

Table 1

Mean (Standard Deviation) [±95% Confidence Limits] Test Scores (as percentages) as a Function of Test Interval, Group, and Lecture Fluency in Experiment 1

	Lecture Fluency		
Test Interval	Fluent	Disfluent	Total
	Control Group		
Immediate Test	62.72 *(17.07)*	55.06 *(11.92)*	58.89 *(12.47)*
	[57.50, 67.93]	[50.28, 59.84]	[54.86,62.92]
Delayed Test	45.68 *(14.52)*	46.42 *(10.70)*	46.05 *(10.09)*
	[40.71, 50.65]	[40.96, 51.88]	[41.89, 50.21]
Total	54.20 *(12.11)*	50.74 *(8.79)*	52.47 *(9.03)*
	[50.02, 58.38]	[46.58, 54.91]	[48.99, 55.95]
	Annotation Group		
Immediate Test	61.38 *(13.95)*	66.43 *(12.66)*	63.91 *(11.13)*
	[56.35, 66.41]	[61.83, 71.05]	[60.02, 67.80]
Delayed Test	47.82 *(12.73)*	49.89 *(14.10)*	48.85 *(10.17)*
	[43.02, 52.61]	[44.62, 55.15]	[44.84, 52.87]
Total	54.60 *(11.73)*	58.16 *(9.86)*	56.38 *(9.44)*

	[50.57, 58.63]	[54.14, 62.18]	[53.02, 59.74]
	NS Group		
Immediate Test	70.86 *(12.56)*	71.11 *(10.29)*	70.99 *(8.66)*
	[65.65, 76.08]	[66.33, 75.89]	[66.95, 75.02]
Delayed Test	56.05 *(11.73)*	59.26 *(16.04)*	57.65 *(11.94)*
	[51.08, 61.02]	[53.80, 64.72]	[63.78, 72.33]
Total	63.46 *(10.56)*	65.19 *(10.99)*	64.32 *(8.85)*
	[59.28, 67.64]	[61.02, 69.35]	[60.84, 67.80]
	NF Group		
Immediate Test	68.33 *(9.48)*	67.78 *(15.05)*	68.05 *(9.42)*
	[62.81, 73.86]	[62.71, 72.85]	[63.78, 72.33]
Delayed Test	57.78 *(13.00)*	55.28 *(15.91)*	56.53 *(11.40)*
	[52.50, 63.05]	[49.49, 61.06]	[52.12, 60.94]
Total	63.06 *(8.79)*	61.53 *(13.83)*	62.29 *(9.10)*
	[58.62, 67.49]	[57.11, 65.95]	[58.60, 65.98]

Note: NS = notetaking (self), NF = notetaking (friends).

Table 2

Mean (Standard Deviation) [±95% Confidence Limits] Confidence Ratings (0-100) for Test Answers as a Function of Test Interval, Group, and Lecture Fluency in Experiment 1

	Lecture Fluency		
Test Interval	Fluent	Disfluent	Total
	Control Group		
Immediate Test	63.62 *(18.86)*	63.27 *(17.11)*	63.45 *(17.03)*
	[57.32, 69.92]	[57.44, 69.10]	[57.85, 69.05]
Delayed Test	45.30 *(19.94)*	43.38 *(19.79)*	44.34 *(18.98)*
	[38.15, 52.46]	[35.81, 50.94]	[37.48, 51.20]
Total	54.46 *(17.52)*	53.32 *(17.39)*	53.89 *(16.83)*
	[48.44, 60.49]	[47.09, 59.55]	[48.04, 59.75]
	Annotation Group		
Immediate Test	66.04 *(11.96)*	59.89 *(12.56)*	62.96 *(10.45)*
	[59.96, 72.11]	[54.26, 65.51]	[57.56, 68.37]
Delayed Test	48.76 *(17.67)*	49.95 *(19.33)*	49.36 *(16.57)*
	[41.86, 55.66]	[42.66, 57.25]	[42.74, 55.98]
Total	57.40 *(12.62)*	54.92 *(15.39)*	56.16 *(12.85)*
	[51.59, 63.21]	[48.91, 60.93]	[50.51, 61.81]
	NS Group		
Immediate Test	70.14 *(16.55)*	65.73 *(12.85)*	67.93 *(13.33)*
	[63.84, 76.44]	[59.90, 71.56]	[62.33, 73.53]
Delayed Test	50.39 *(15.67)*	51.21 *(20.07)*	50.80 *(16.62)*
	[43.24, 57.54]	[43.65, 58.78]	[43.94, 57.66]
Total	60.26 *(14.44)*	58.47 *(14.88)*	59.37 *(14.14)*
	[54.24, 66.29]	[52.24, 64.70]	[53.52, 65.22]
	NF Group		
Immediate Test	64.76 *(18.42)*	62.92 *(18.56)*	63.84 *(17.62)*
	[57.78, 71.74]	[56.46, 69.38]	[57.63, 70.04]
Delayed Test	48.05 *(21.79)*	43.36 *(20.17)*	45.71 *(19.99)*
	[40.13, 55.97]	[34.98, 51.75]	[38.11, 53.31]
Total	56.40 *(18.61)*	53.14 *(17.78)*	54.77 *(17.65)*
	[49.73, 63.08]	[46.24, 60.04]	[48.29, 61.25]

Note: NS = notetaking (self), NF = notetaking (friends).

Table 3

Mean (Standard Deviation) [±95% Confidence Limits] Number of Words (Word-Count) as a Function of Lecture Fluency and Group in Experiment 1

	Lecture Fluency		
Group	Fluent	Disfluent	Total
Annotation	55.31 *(18.79)* [37.61 73.01]	60.44 *(21.18)* [42.06, 78.27]	57.88 *(16.68)* [41.70, 74.05]
NS	108.93 *(39.94)* [95.30, 124.06]	110.33 *(37.40)* [96.61, 124.06]	109.63 *(33.39)* [97.18, 122.08]
NF	128.22 *(38.28)* [114.60, 141.85]	123.93 *(38.74)* [110.20, 137.65]	126.07 *(36.72)* [113.62, 138.53]

Note: NS = notetaking (self), NF = notetaking (friends).

Table 4

Mean (Standard Deviation) [±95% Confidence Limits] Number of Test Answers (/30 questions) Written in the Notes/Annotations (Correspondence) as a Function of Lecture Fluency and Group in Experiment 1

	Lecture Fluency		
Group	Fluent	Disfluent	Total
NS	18.76 *(5.36)* [16.72, 20.80]	18.57 *(6.79)* [15.93, 21.22]	18.67 *(5.38)* [16.63, 20.71]
NF	19.84 *(3.61)* [17.70, 21.98]	19.84 *(4.94)* [17.06, 22.62]	19.84 *(3.59)* [17.70, 21.99]

Note: NS = notetaking (self), NF = notetaking (friends).

Table 5

Mean (Standard Deviation) [±95% Confidence Limits] Test Scores (as percentages) at Immediate Testing as a Function of Lecture Pace, Information Type, and Group in Experiment 2

	Lecture Pace		
Lecture Material	Faster	Regular	Total
	Control Group		
Fact	62.32 *(18.84)*	65.58 *(22.54)*	63.95 *(15.64)*
	[54.51, 70.13]	[57.03, 74.13]	[57.54, 70.36]
Concept	39.86 *(24.89)*	43.84 *(24.93)*	41.85 *(21.79)*
	[29.47, 50.24]	[34.13, 53.56]	[33.78, 49.92]
Total	51.09 *(18.18)*	54.71 *(18.55)*	52.90 *(15.17)*
	[43.69, 58.49]	[47.53, 61.89]	[47.10, 58.70]
	Annotation Group		
Fact	56.58 *(19.19)*	55.26 *(20.78)*	55.92 *(14.30)*
	[47.99, 65.17]	[45.86, 64.67]	[48.87, 62.98]
Concept	33.33 *(28.84)*	42.98 *(22.51)*	38.16 *(18.54)*
	[21.91, 44.76]	[32.29, 53.67]	[29.28, 47.04]
Total	44.96 *(17.77)*	49.12 *(17.43)*	47.04 *(11.99)*
	[36.82, 53.10]	[41.22, 57.02]	[40.66, 53.42]
	NT Group		
Fact	70.00 *(18.17)*	72.92 *(13.11)*	71.46 *(13.30)*
	[61.62, 78.38]	[63.75, 82.09]	[64.58, 78.33]
Concept	42.92 *(24.96)*	47.92 *(21.14)*	45.42 *(18.59)*
	[31.78, 54.06]	[37.50, 58.34]	[36.76, 54.07]
Total	56.46 *(17.06)*	60.42 *(13.13)*	58.44 *(12.94)*
	[48.52, 64.39]	[52.72, 68.12]	[52.22, 64.66]

Note: NT = Notetaking.

Table 6

Mean [±95% Confidence Limits] Test Scores (as percentages) at Delayed Testing as a Function of Lecture Pace, Information Type, and Group in Experiment 2

	Lecture Pace		
Lecture Material	Faster	Regular	Total
	Control Group		
Fact	36.23 *(16.02)* [28.01, 44.45]	39.54 *(18.19)* [32.64, 46.35]	37.86 *(12.77)* [31.82, 43.90]
Concept	27.54 *(18.18)* [18.90, 36.17]	24.28 *(25.05)* [15.94, 32.60]	25.91 *(16.14)* [19.34, 32.47]
Total	31.88 *(11.69)* [25.54, 38.22]	31.88 *(16.46)* [26.14, 37.63]	31.88 *(11.12)* [26.97, 36.79]
	Annotation Group		
Fact	39.47 *(18.49)* [30.43, 48.52]	33.33 *(14.19)* [25.79, 25.79]	36.40 *(13.66)* [29.76, 47.04]
Concept	19.30 *(17.52)* [9.80, 28.80]	23.25 *(17.44)* [14.08, 32.41]	21.27 *(14.58)* [14.05, 28.50]
Total	29.39 *(15.91)* [22.41, 36.36]	28.29 *(11.23)* [21.97, 34.61]	28.84 *(11.33)* [23.44, 34.24]
	NT Group		
Fact	49.58 *(20.24)* [40.77, 58.40]	45.42 *(19.70)* [38.07, 52.77]	47.50 *(16.70)* [41.02, 53.98]
Concept	28.33 *(23.42)* [19.08, 37.59]	43.75 *(15.42)* [34.82, 52.68]	36.04 *(16.35)* [29.00, 43.09]
Total	38.96 *(14.78)* [32.16, 45.76]	44.58 *(12.52)* [38.42, 50.75]	41.77 *(11.40)* [36.51, 47.04]

Note: NT = Notetaking.

Paper 2 Overview

Paper 1 demonstrated that longhand notetaking is more beneficial for encoding during lectures than slide annotation or passive observation, regardless of the fluency of the lecture delivery, the speed of the lecture delivery or whether the lecture information pertains to facts or concepts.

For our second paper, we planned to replicate these findings using different material, but we also included three extra encoding conditions, to investigate whether longhand notetaking is optimal, or whether further intructions can be given to assist students during encoding.

Firstly, we added a verbatim notetaking group, who were given blank paper and were instructed to write down as much of the lecture content as they could, exactly as it appeared in the recorded lecture. This group were included in order to identify if all notetaking is equally beneficial, or if this type of notetaking is less advantageous to students, as it is not generative. Bui, Myerson and Hale (2013) found that verbatim notetaking led to very good performance on immediate testing, but, following delay, showed significantly poorer performance than the notetakers instructed to organise their notes. Therefore, we predicted that this group in our experiment would show good performance at immediate testing, but reduced performance following delay.

We also included two further groups who did not make notes during the lecture, but were provided with interspersed pauses throughout the material in which to generate the three key points from the previous lecture section from memory. One group were not provided feedback for their answers but a second group were shown a slide with the correct key points, but were instructed not to change their answers. These groups were included to investigate whether a form of retrieval practice, which has been repeatedly shown to be beneficial as a restudy aid (see Dunlosky, Rawson, Marsh, Nathan & Willingham, 2013; Roediger & Karpicke, 2006; Rowland, 2014) can be integrated into a lecture to form an encoding strategy. We predicated that these groups, particularly the group receiving feedback, would show the best performance amongst all our groups, outperforming controls (who passively observed), slide annotators, verbatim notetakers and regular notetakers.

In order to investigate further impacts of our encoding strategies, we also took measures of retrospective confidence, as in Paper 1, judgements-of-learning, task unrelated thoughts (TUTs) and enjoyment. However, the scale used to measure enjoyment was deemed retrospectively unclear, as participants could have been confused by the question and answered using the reverse scale. Therefore, this measure was dropped. Information relating to all other measures not pertaining to performance can be found in the supplementary findings.

Paper 2 – Activities that Enhance Learning During Lectures

Manuscript Submitted to Contemporary Educational Psychology: 06/07/18

Abstract

Coria and Higham (2018a) showed that longhand notetaking during lectures leads to improved memory of lecture material compared to annotating lecture slide handouts (LSHs) or passive observation. Here, we replicate those results and investigate ways to enhance the benefits of notetaking. Participants watched a lecture on MRSA in one of six encoding groups: passive observation, LSH annotation, verbatim notetaking (copying slide information), regular notetaking (taking notes as if in normal lecture), retrieval practice (self-generating key lecture points during specified pauses), and retrieval practice with immediate feedback. On both immediate and delayed memory tests, both retrieval-practice groups scored the highest, with, contrary to most research, no further effect of feedback. Regular notetaking led to good performance on delayed tests, whereas verbatim notetaking led to good immediate test results, but poor results on the delayed test. LSH annotators and passive observation groups scored the lowest and did not differ.

Our findings are consistent with the *New Theory of Disuse* and suggest that notetaking in the form of retrieval practice is a highly effective learning strategy that could easily be implemented in lectures.

Keywords: *Lecture Activities, Encoding Function, Notetaking, Retrieval Practice, Feedback*

Activities that Enhance Learning During Lectures

Modern technology has made available a number of different methods for presenting material to students in lectures, some of which are believed to promote learning. For example, most university instructors have switched from the chalkboard or projector as the primary delivery mode of lecture content to electronic presentation software, such as Microsoft PowerPoint or Apple Keynote. These software packages are typically used to divide lecture content into a series of consecutive "slides," which often explicitly state key points and other important information. Copies of these slides can be (and frequently are) distributed to students to enable them to have personal copies during lectures. The consequence is that the activities that students engage in within lectures has changed dramatically over the past few decades. A particularly salient example of this change is that longhand notetaking on blank paper has largely been replaced with annotating either paper or digital versions of lecture slide handouts (LSHs).

LSHs vs Notetaking: The Storage Function

There is a widespread belief that distributing LSHs to students constitutes "good practice," evinced by the fact that universities are increasingly mandating that these slides are made available to students, even in advance of lectures. For example, instructors may be required to post them on virtual learning environments such as Blackboard for students to download. One potential learning benefit of LSHs is that they often provide a complete and accurate record of lecture content that can later be used for revision purposes. In the context of research on longhand notetaking, Di Vesta and Gray (1972) described this type of record-keeping as an "external storage mechanism" (p.

8). Kiewra (1985) found that following a revision session, participants with the most complete longhand notes showed the best performance at testing. Presumably, LSHs could be a more accurate and possibly a more complete record of the lecture material than even the best students' longhand notes. Therefore, in terms of this external storage mechanism, LSHs could potentially improve test performance even more than complete longhand notes.

LSHs vs Notetaking: The Encoding Function

In addition to the external storage function, Di Vesta and Gray (1972) identified the *encoding function* of notetaking. That is, it is possible to investigate the effect of notetaking on the encoding of lecture information independently from external storage effects by, for example, testing students on lecture content without providing a revision opportunity. Kobayashi (2005) found in a meta-analysis of the encoding function of notetaking that, compared to passive observation, students who take notes tended to perform modestly better on memory tests, even if there has been no opportunity for revision, and particularly if learning is tested using a cued- or free-recall format.

The same potential encoding function may exist for LSHs as well. For example, LSHs may promote lecture comprehension by providing structure to the lecture and the means to "get back on track" after lapses of attention. They also remove the requirement to write down key information, which, if left unchecked, could strain attentional resources, particularly if the lecture pace is fast. In other words, LSHs may facilitate learning by promoting *efficient encoding* (Marsh & Sink, 2010) of lecture material, with less distraction and less work, thereby making the lecture experience more enjoyable. In this vein, a survey conducted by Coria and Higham (2018a) found that 61% of undergraduate students rated lectures as more difficult without the presence of an LSH.

Perhaps for this reason, research has unanimously demonstrated that LSHs are very popular amongst students (e.g., Babb & Ross, 2009; Coria & Higham, 2018a, Daniels, 1999; Douglas, Douglas & Barnes, 2006; Marsh & Sink, 2010)

However, whilst it is true that notetaking can be cognitively demanding (Piolat, Olive, & Kellogg, 2005) with those demands mostly alleviated by LSHs, is removing these difficulties necessarily positive? Intuition suggests that it is. However, recent research suggests that both students and educators possess a number of erroneous intuitions about what is good for learning (e.g., see Kornell, Rhodes, Castell & Tauber, 2011; Rhodes & Castell, 2008; Serra & Dunlosky, 2010). For example, many people believe that *easy* learning is associated with *long-term* retention in memory, whereas in fact, the opposite is usually true (e.g., Bjork, Dunlosky, & Kornell, 2013; Besken & Mulligan, 2013; Miele, Finn, & Molden, 2011). That is, often it is the case that the harder it is to encode information, the longer it will last in memory, a phenomenon encapsulated by Bjork's (1994) principle of *desirable difficulty.* The fact that notetaking has been shown to be difficult and to facilitate later test performance (Kobayashi, 2005) suggests that it should probably be considered a desirable difficulty.

Thus, one potential concern about LSHs is that they largely replace the desirably difficult process of longhand notetaking with more passive encoding processes such as reading and/or modest amounts of LSH annotation. Therefore, when considering the encoding function separately from the storage function, the question becomes, do the potential benefits of LSHs in terms of lecture comprehension and maintaining attention described earlier outweigh the potential encoding benefits of notetaking? Despite the ubiquity of LSHs and the critical importance of this question, there have only been a few studies addressing it, to which we now turn.

Research Comparing the Encoding Function of LSHs vs Notetaking

Marsh and Sink (2010) conducted two experiments directly comparing test performance with LSHs present vs absent. Participants were permitted to revise the material in some conditions, but the no-revision conditions of their first experiment isolated the encoding function. In those conditions, participants viewed a pre-recorded lecture either with a handout present (i.e., they were given blank paper and a pen for notetaking plus a LSH to annotate) or with no handout (i.e., they were provided only with blank paper and a pen for notetaking). They were then tested either immediately or after a delay but were given no opportunity to revise the material. Marsh and Sink found that participants in the handout-present and handout-absent conditions performed equally regardless of the length of the retention interval.

At first blush, the equal test performance between the handout-present vs handout-absent groups would suggest that handouts are ineffectual if the encoding function is isolated. However, this conclusion is compromised by the fact that notetaking also varied between the groups. That is, participants in the handout-absent group wrote over twice as many notes as participants in the handout-present group. Thus, it is not clear from Marsh and Sink's (2010) results whether the equal test performance between these groups was due to both handouts and notetaking being ineffectual, or whether each had beneficial effects specific to one group only. That is, the handout-present group may have benefitted from handouts but not from notetaking, whereas the opposite was true in the handout-absent group. If so, this scenario could also potentially lead to comparable performance between the groups.

Coria and Higham (2018a) also compared the encoding function of longhand notetaking and LSH annotation, but avoided the interpretive problems present in Marsh and Sink's (2010) research by adding a control group. Specifically, in this study, participants first watched two pre-recorded, 10-min lectures on medical topics. Prior to the lectures, they were provided with either blank paper and a pen for notetaking (one group took notes for themselves; another for an absent friend), printed LSHs and a pen to annotate the slides, or no materials. The participants given no materials acted as the control group and they were instructed to just watch and listen to the lecture without a LSH and without taking notes. By including this "passive observation" control group, it was possible to separate the potential beneficial effects of notetaking vs LSH annotation. Coria and Higham found that, compared to controls, participants who wrote longhand notes (either for themselves or an absent friend) had better immediate- and delayed-test performance. In contrast, LSH annotators did not perform better at either immediate or delayed testing than participants in the control group. Thus, the inclusion of the necessary control group allowed Coria and Higham to conclude that notetaking conferred benefits to learning whereas LSHs conferred none.

In addition to these laboratory-based studies, some more naturalistic research has been conducted comparing LSHs and notetaking within real academic lectures. Murphy and Cross (2002) and Weatherley, Grabe, and Arthur (2003) both found an advantage of notetaking over LSHs. Conversely, Babb and Ross (2009) observed that there was no difference between a group of students given LSHs before lectures (who took few longhand notes) vs another given the LSHs afterwards (who took many longhand notes). However, the results of these naturalistic studies are difficult to interpret because of extraneous, confounding variables. For example, the data for two groups compared in Babb and Ross's study were collected in different semesters, in different courses, and these courses had different content and lecture length. Any of these uncontrolled variables could have affected the results.

Regardless of these interpretive problems, it is clear that the literature as a whole is equivocal about whether it is preferable for students to take longhand notes during lectures or to encode the lecture given LSHs to annotate. Consequently, the current research was conducted to contribute to this debate by both seeking to replicate and extend our earlier results on the beneficial effects of notetaking over LSH annotation (Coria & Higham, 2018a) and to investigate other lecture activities that might enhance learning even further.

Verbatim vs Self-Generated Notetaking

As noted earlier, Kobayashi (2005) found that notetaking was beneficial for learning even when the encoding function was examined in isolation. However, he described that benefit as a "positive but modest effect" (p. 253). A potential reason for the effect only being modest is that there may be a number of factors that moderate the benefits of notetaking. Various studies have examined some of these moderators. For example, Peverly and Sumowski (2012) investigated factors such as transcription fluency and the quality of notes. They found that the former predicted the latter, and that quality of notes predicted test scores. Peverly, Brobst, Graham, and Shaw (2003) investigated the role of inference generation during notetaking and its impact on test scores. Their results suggested that inferences generated during notetaking were particularly helpful on test questions that required inferences instead of rote memory.

In the current research, we compared verbatim notetaking with students' more natural tendency to mix verbatim notes and self-generated notes. Verbatim notetaking is frequently dismissed as low utility. For example, Mueller and Oppenheimer (2014) found that notetaking on a laptop was less beneficial to memory than longhand notetaking. They attributed this difference to laptop users simply transcribing lecture-slide material given that most modern students can type faster than they can write. Longhand note takers, on the other hand, had a greater tendency to paraphrase the information in their own words and focus on key points. In the same vein, Bretzing and Kuhlhavy (1979) manipulated how participants took notes and found that, in the absence of revision, the verbatim group showed lower test scores than both the summarizing and paraphrasing groups. Overall, this research suggests that there is an advantage to self-generating notes rather than just copying them. This conclusion is bolstered by Armbruster (2000) who concluded that that the more generative the notetaking activity, the more likely learning is to occur.

However, is verbatim notetaking always worthless for learning? Recent research into the *production effect* (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010) suggests that it may not be. This research has shown that, compared to reading, “producing” information can be beneficial to memory. Although most research into the production effect has focused on the vocal reproduction of information (i.e., saying it out loud), Mama and Icht (2016) found an even greater advantage for writing information compared to vocalizing it. These results imply that copying slide information could benefit memory more than simply reading information on an LSH.

A recent study by Bui, Myerson, and Hale (2013) supports this conclusion. They found that verbatim notetaking led to very good performance on an immediate test. In fact, it was even better than the performance of a group of students instructed to paraphrase and organize their notes! This advantage was short lived however; it was not just attenuated, but *reversed,* on a delayed test taken 24-hours later, making these results more in line with those discussed previously. Nonetheless, Bui et al.'s results suggest that verbatim notetaking can confer some memorial advantages, at least in the short term. Additionally, verbatim notetaking may have other advantages. For example, compared to paraphrased notetaking, it may reduce the demands on attentional resources that derive from the process of selecting and paraphrasing important information (Cohn, Cohn, & Bradley, 1995). These reduced demands may, in turn, mean that students' notes are more complete, which may be beneficial to encoding. In this vein, research by Einstein, Morris, and Smith (1985) found that there was only a 5% chance of important information being later recalled if it was not written during the encoding phase.

Retrieval Practice

Although most of the research discussed earlier suggests that self-generating or paraphrasing notes is more beneficial to memory than copying lecture-slide information, this does not necessarily make paraphrased notetaking an optimal strategy for learning. There are a variety of activities known to be effective for learning that could be instantiated in lectures by educators which could potentially further improve lecture encoding. One such activity is *retrieval practice*.

Countless studies conducted over many decades have demonstrated the memorial advantages of retrieval practice (e.g., Karpicke & Blunt, 2011; McDaniel, Anderson, Derbish, & Morissette, 2007; Roediger & Butler, 2011; see Roediger & Karpicke, 2006 and Rowland, 2014 for reviews). In short, after first learning some material, practice at retrieving that material after an interval (e.g., via practice quizzes) leads to significantly better memory on both short- and long-term memory tests compared to restudying that same material. The research on the efficacy of retrieval practice is so overwhelming that

Dunlosky, Rawson, Marsh, Nathan, and Willingham (2013) considered it to be one of the highest utility learning techniques of those they reviewed. Furthermore, retrieval practice is known to facilitate transfer of previous knowledge to novel contexts (Carpenter, 2012), clearly a desirable goal for educators.

Retrieval practice has mostly been discussed as a strategy that can be adopted after initial learning is complete (e.g., as an effective exam revision activity). However, some other research has found that it may be beneficial even during encoding, which makes it something that could potentially be incorporated into lectures. Di Vesta and Gray (1972) found that testing during an interval following a 5-min recorded communication improved later test performance compared to no testing. However, their design confounded exposure to the material with testing. That is, the tested groups may have performed better on the final test not because of retrieval practice per se, but because they had an additional opportunity to learn the communication on the initial test. More recently, however, exposure opportunities have been better controlled. For example, Lyle and Crawford (2011) found that testing students immediately following a lecture produced significantly higher exam scores than those given time to review the content.

More recently, Szpunar, Khan, and Schachter (2013, Experiment 2) also controlled for re-exposure effects by comparing test performance for a group that was tested four times throughout a lecture (interpolated testing) to another that was given opportunities to restudy the material four times (interpolated restudying). They found substantial benefits on a final test for the tested group. Healy, Jones, Lalchandani, and Tack (2017) compared final test performance between a group given several short, interpolated quizzes during learning and one given a single, longer quiz at the end of the lecture (prior to the final test). They found superior performance for the interpolated-quiz group. Indeed, the benefits in the interpolated-quiz group were observed even for questions that were not tested, suggesting that repeated testing serves to facilitate learner engagement (see also Szpunar, Khan & Schachter). However, what is not clear at present is whether repeated retrieval practice during lectures is more or less advantageous compared to normal notetaking. We attempted to answer this question in the current research by directly comparing the memory performance of groups of participants who took either verbatim or paraphrased notes during a lecture with a group who took no notes except during specified pauses in the lecture. During these pauses, participants were instructed to practice retrieving key lecture information and hence these groups were similar to the interpolated testing groups described above.

Exactly how normal notetaking and notetaking in the form of retrieval practice will compare may depend in part of how accurate participants are at retrieving the correct information. If participants retrieve nothing but erroneous information (i.e., errors of *commission*), then retrieval practice could backfire. Specifically, this erroneous information could become strengthened in memory by retrieval practice, thereby increasing the likelihood that it is reproduced on later tests, worsening performance rather than enhancing it. Alternatively, there could be total retrieval failure (i.e., errors of *omission*), in which case key information might benefit more from a restudy opportunity rather than retrieval practice (cf. the bifurcation distribution model – Kornell, Bjork, & Garcia, 2011).

However, research has generally shown that corrective feedback is largely preventative of these undesirable outcomes (e.g., Butler, Karpicke, & Roediger, 2007). As long as students receive corrective feedback, making errors during retrieval can sometimes *benefit* later memory performance. This is particularly true of high-confidence errors, which are more likely to be corrected on later tests than low-confidence errors, an effect dubbed *hypercorrection* (e.g., Butterfield & Metcalfe, 2001). However, sometimes, even items that produce low-confidence errors (i.e., incorrect guesses) are more likely to be correctly responded to on later tests than items that are merely studied (e.g., Kornell, Hays, & Bjork, 2009; Potts & Shanks, 2014; see Griffiths & Higham, 2018, for a particularly salient example of good low-confidence error correction). For these reasons, we also included a retrieval-practice group that was given corrective feedback on their retrieval attempts.

Experiment 1

The current research was designed to replicate Coria and Higham's (2018a) demonstration of the memorial benefits of notetaking while also gaining some control over the type of notes that students took during a simulated lecture. At the same time, we also sought to identify other, potentially more effective, activities that educators could instantiate in their lectures to encourage students to better encode the lecture material. Altogether, we tested six groups of participants. The procedure for three groups (control group, annotation group and regular-notes [RN] group) was similar to that used in Coria and Higham, discussed earlier. In addition, we included three new groups to elucidate the type of notetaking that produces memorial benefits and compared the size of those benefits to groups engaged in retrieval practice. These three additional groups were the verbatim-notes (VN) group, the retrieval-practice (RP) group and the retrieval-practice-with-feedback (RPF) group.

Participants in the VN group, as the name suggests, were required to write as much as much of the information from the lecture slides as possible *exactly* as it was written.

Due to the fact that this group was not writing anything in their own words, any benefit in this group over passive observation could be attributed to something akin to the production effect. Participants in the RP group were asked not to write anything until specific pauses in the lecture, at which point, they were asked to retrieve three key points from the just-completed lecture section. Participants in the RPF group also engaged in retrieval practice in the same way as the RP group, but following retrieval practice, they were shown a slide listing the three key points that participants were expected to retrieve. Despite the feedback, participants were not allowed to change their responses. Any advantage in this group over the RP group could be directly attributed to the effect of feedback.

Participants were assigned to one of six encoding groups and viewed a lecture on a medical topic – the hospital super bug MRSA. Participants were tested on the lecture information both immediately following the lecture and after a week-long delay. On both tests, participants answered both multiple choice and short-answer (cued-recall) questions.

Critically, because we wanted to focus exclusively on encoding effects unaffected by storage effects, participants were not permitted to engage in any form of review of the lecture material prior to testing.

Method

Participants. An a priori power analysis using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) was conducted under the assumption that our key findings would relate to a difference in the participants' test scores across the encoding groups, and that we would be making between-subjects, pairwise comparisons across these groups. The analysis was conducted with power (1 - β) set at .80 to determine how many participants would be needed (per group) to detect an effect size of d = 1 (equivalent to η_p^2 = .20). This effect size was selected based on an analogous effect size for notetaking in Coria and Higham (2018a). Based on these parameters, G*Power indicated that a sample of 17 participants per group was the minimum requisite.

In total, 147 undergraduate students from the University of Southampton (29 males, 118 females, *M* age = 20.81, *SD* age = 2.97) took part in this research in exchange for course credit. The participants were assigned to an encoding group based on the session that they signed up to attend. All 147 participants attended the first session, but five participants were excluded (two in the VN group, one in the RP group, and two in the RPF group) for not complying with the experimental procedure and a further participant (in the VN group) was excluded after indicating that she had previous knowledge on the lecture topic. The remaining 141 participants were assigned to six encoding groups (explained in more detail later) as follows: control = 34, annotation = 22, VN = 19, RN = 23, RP = 21, and RPF = 22. In Phase 2, 17 of the 141 participants who were present for Phase 1 did not return, leaving the following sample sizes for Phase 2: control = 24, annotation = 20, VN = 19, RN = 21, RP = 18, RPF = 22.

Design and materials. The design involved six independent encoding groups (control, annotation, VN, RN, RP, and RPF) tested at two delays (immediate vs 1-week). Our primary interest was in test accuracy (based on both multiple-choice and cued-recall questions). We also analyzed the content of the notes, as explained in more detail later. In addition to these measures of primary interest, we also included a number of exploratory dependent variables such as retrospective confidence, lecture enjoyment, aggregate judgments of learning (aJOLs), and task unrelated thoughts (TUTs). The results obtained from these measures were largely inconclusive, so we report only test accuracy and the results of note-content analyses. Limiting our focus in this way also served the interests of keeping our report briefer and more focused. For readers interested in the results from these exploratory dependent variables, please contact the authors.

The experiment took place over two separate sessions: Phase 1, in which the participant viewed the lecture and completed an immediate memory test, and Phase 2, which ensued seven days later, in which delayed memory testing occurred. The encoding groups differed in the activity that they engaged in whilst watching the lecture. All of the activities and instructions given to the different groups are summarized in Table 1. The control group was given no lecture materials and simply instructed to watch the lecture. The annotation group was given a pen and a printed copy of the slides to guide them through the lecture. They were instructed to listen to the lecture and annotate the handout as if in a real lecture. The VN and RN groups were both provided with a pen and blank sheets of paper. The only difference between these two groups was that the VN group was instructed to copy as much slide content as possible whereas the RN group was instructed to take notes as if they were in a real lecture. The RP and RPF group were both provided with a pen and paper, except, unlike the notetaking groups (RN and VN), the pages were not blank. Instead, there was 10 boxes printed over two sheets corresponding to the 10 sections of the lecture, each with the labels “1,” “2,” and “3,” to write three key points. Participants were instructed to write nothing until prompted by the voiceover, so no notetaking occurred except in the allocated 1-min, retrieval-practice pauses. Following the key-points retrieval practice, the RPF group only was provided with feedback. In particular, they were shown a box on the screen at the front of the lecture (like that printed on their lecture materials, with the labels "1," "2," and "3") with the three key points filled in for that section.

Two tests, consisting of 30 questions in total (one question relating to each of the 30 key points in the lecture) were created. Participants took one test at the end of Phase 1 (immediate test) and a second, different test in Phase 2 one-week later (delayed test). The order of the tests was counterbalanced across participants. Each test consisted of 15 questions relating to 15 of the key points covered in the lecture, with at least one question from each of the ten sections. Ten of the questions on each test were multiple-choice (each with five options) and five were cued-recall questions. The questions on the test appeared in the order in which the relevant material occurred in the lecture, with cued-recall and multiple-choice questions interspersed throughout. All participants wrote the same tests with the same orders of questions; only the order of the tests was counterbalanced.

Procedure. Both phases of the experiment took place in the same, medium-sized lecture theatre (maximum capacity = 75) and the experiment was conducted in groups of between four and 16 participants at one time, who were all in the same encoding group.

Phase 1 - lecture. On arrival, participants were seated in the front three rows of the lecture theatre, at least two seats apart from each other. They were given a screening questionnaire in which they were asked to provide their age, gender, year of study, and whether they had previously studied or been affected by the lecture topic. Once all the participants had completed the questionnaire and were ready to proceed, they received the verbal instructions and material corresponding to their encoding group outlined in Table 1.

All participants also received a questionnaire on A4 paper entitled "Task-Unrelated Thoughts" with ten left-justified boxes running vertically down the page. The participants were verbally instructed that at ten points during the lecture, there would be a pause and they would be asked to provide a value between zero ("not at all") and nine ("consistently throughout the entire section") to indicate how much they had been experiencing thoughts that were not related to the lecture in the section prior to the pause. They were informed that during the lecture, these instructions would be repeated each time they were required to complete this task.

The lecture, which was about the hospital superbug MRSA, was pre-recorded and played on a large screen at the front of the lecture theatre, with the audio projected from speakers both at the front and back of the lecture theatre. It consisted of Apple Keynote lecture slides designed to mimic those seen in typical university classrooms. All slides were written in bullet-point format and some slides contained supplementary images, but the main lecture content was represented by the text on the slides. The slides were accompanied by a pre-recorded audio voiceover conveying information in the style of a lecturer. The lecture was pre-recorded to ensure that participants in all sessions received the same material, delivered at the same pace and with the same voice intonation (i.e., to maintain experimental control).

The whole lecture for all groups consisted of 10 blocks each consisting of a lecture section plus additional activities that varied according to group (see Figure 1). Each lecture section was approximately 4-min in length and consisting of three slides of content. Each section related to a different aspect of the topic (such as *mechanisms of infection* or *prevalence*) and each section included three main, high-importance messages, or *key points* as they were referred to throughout the experiment.

Information pertaining to each key point was included in both the slides and audio voiceover although the wording was not necessarily identical. This was done to mimic a true lecture experience.

The additional activity for each of the 10 blocks for the control, annotation, VN, and RN groups was 30 seconds to complete the TUT questionnaire (10 x 4.5-min = 45-min total). Specifically, at the end of each lecture section within each block, the voiceover prompted participants to use their TUT questionnaire to provide a rating of how much they had been experiencing task-unrelated thoughts during the just-completed lecture section. For each block in the RP group, following the lecture section and completion of the TUT questionnaire, an additional 1-min was added for participants to practice retrieving the three key points corresponding to that section (10 x 5.5-min = 55-min total). Finally, in addition to the lecture section, TUT questionnaire, and retrieval practice, the RPF group was given 30 secs to view a feedback slide displaying the key points that should have been retrieved (10 x 6-min = 60-min total). Thus, although the lecture content was identical across all the encoding groups, the lecture delivery time ranged from 45 – 60 min depending on group.

Immediately following the lecture, participants were given a questionnaire where they were asked to provide one numerical value (0-9) to indicate their enjoyment of the lecture and ten aJOLs, one for each of the ten sections. For the aJOLs, participants were asked to estimate the likelihood (as a percentage) that they would be able to answer questions correctly from each section if tested. The sheet for the aJOLs listed the titles of the 10 lecture sections in the order that they had occurred in the lecture, to help participants remember which section the aJOLs pertained to. All groups spent exactly 5-

min completing the questionnaire to ensure a constant retention interval between the end of the lecture and testing.

After providing aJOL judgments, participants engaged in a 10-min filler task. For this task, participants were asked to write down (on a sheet of provided paper), as many alternative (i.e., unintended) uses for eight common household items, such as a candle. Again, this was kept to exactly 10-min per group to ensure a 15-min interval between the end of the lecture and testing (5-min for the questionnaire and 10-min for the filler task).

***Phase 1 – test 1*.** On termination of the filler task, participants were tested on the content of the lecture. Test 1 (and the test completed after a delay – see later) had two cover pages of instructions on how it should be completed (accompanied by verbal instructions). For the multiple-choice questions, participants were instructed to rank the five answer options that appeared horizontally in a table, with “1” vs “5” representing the answer they believed was the most vs the least likely to be correct, respectively. The answers to each multiple-choice question had been deliberately constructed to vary in plausibility, which had been independently verified by two medical professionals. Participants were also instructed to indicate their confidence (0-100%) that their highest-ranked answer was correct. An example of a correctly completed multiple-choice question (unrelated to the lecture material) was included on the cover page.

For the cued-recall questions, participants were given space to provide an answer and to rate confidence (0-100%) alongside the instruction to guess if unsure (i.e., not to leave blanks). Examples of both multiple-choice and cued-recall questions can be seen in Figure 2. All questions (and correct answers) on the test were explicitly featured within the lecture and participants were not expected to draw on any external information to score full marks.

To encourage optimal performance, there was no time limit on the test, but participants were asked to indicate when they had completed the test so that the experimenter could provide them with a partial debriefing form reminding them to return a week later. The partial debriefing form also instructed them to contact the researchers or consult with their doctor if they had any concerns prior to Phase 2 of the experiment. There was no mention of the aim of the study, nor of a second test in Phase 2 to prevent participants from researching the topic between sessions.

Phase 2 – test 2. After seven days, participants returned to the same lecture theatre and were given a second, different test. Once again, the test was self-paced and consisted of ten multiple-choice questions and five cued-recall questions. As in Phase 1, participants were asked to provide a confidence rating for every cued-recall answer as well as for the highest ranked multiple-choice answer. Finally, before leaving, participants were fully debriefed and thanked for their time. If participants did not return for Phase 2, they were sent full debriefing information by email.

Scoring. For the multiple-choice questions, we assigned 1 point (i.e., correct) if the correct answer was assigned rank 1 and zero (i.e., incorrect) if it was assigned any other rank. The cued-recall questions were scored as either correct (1) or incorrect (0). No partial credit was assigned for answers that were partially correct. As stated above, all test questions and expected correct answers were contained within the lecture material, therefore credit was only awarded in the case of cued-recall questions where answers either entirely or almost entirely matched the correct answer, and ambiguous answers were marked as incorrect. For example, for the question “MRSA is normally asymptomatic, but what happens to a person to cause the onset of symptoms?” the correct answer, as described in the lecture was “a weakened immune system”. The answer “a persons immune system becomes compromised” would also be accepted as would be considered as having an equal meaning to the correct answer, and would score one point, but the answer “they get ill”, although partially true, would be considered too general thus not be accepted and the participant would receive zero points for that cued-recall answer.

The main statistical analyses were based on the mean total test scores per participant which included both multiple-choice and cued-recall questions. To create the total test scores, the scores for each test type were weighted 2:1 in favor of multiple-choice questions because there were 10 multiple-choice questions and five cued-recall questions. Weighted final test scores were used for the main analyses because they most closely resemble the type of scores that students receive on typical university examinations, which often consist of a mixture of multiple-choice and short-answer items. As each question, regardless of its type, related to a single key point, and each key point was considered to be of equal value, we assigned equal value to each question, regardless of question type.

Ethics approval. The study was reviewed and approved by the University of Southampton Ethics Committee and the Research Governance Office. All participants were consenting adults over 18 years old and all gave informed consent for their data to be included in the study and any publications that follow from it. Following completion of the study, participants were fully debriefed.

Results and Discussion

Immediate test accuracy. In order to preserve the significantly increased sample size for immediate testing, immediate and delayed scores were analysed separately.

Table 2 shows mean accuracy (and 95% confidence intervals) on the immediate test as a function of encoding group and test type. The data are divided in Table 2 according to test type (multiple choice vs cued recall) for completeness but were collapsed over test type for purposes of analysis. These collapsed means are shown in the far-right column of Table 2. A one-way, between-subjects Analysis of Variance (ANOVA) with six levels (group: control, annotation, VN, RN, RP, and RPF) conducted on these data revealed a main effect of group, $F(5,135) = 5.60$, $p < .001$, $\eta_p^2 = .17$.

To investigate the main effect of group in more detail, we conducted one-way ANOVAs between the control group and each of the experimental groups. These comparisons revealed no significant differences between the control group and the annotation group, $F < 1$, nor between the control group and the RN group, $F(1,55) = 2.38$, $p = .128$, $\eta^2 = .04$. However, we found a significant difference between the control group and the VN group, $F(1,51) = 7.57$, $p = .008$, $\eta^2 = .13$, $d = .77$, the control group and the RP group, $F(1,53) = 12.24$, $p = .001$, $\eta^2 = .19$, $d = .96$, and the control group and the RPF group, $F(1,54) = 11.52$, $p = .001$, $\eta^2 = .18$, $d = .92$.

Delayed test accuracy. Table 3 shows mean accuracy (and 95% confidence intervals) on the delayed test as a function of encoding group and test type. Another one-way, between-subjects ANOVA with six levels (group: control, annotation, VN, RN, RP, and RPF) was conducted on overall delayed-test accuracy. It also revealed a main effect of group, $F(5,118) = 8.30$, $p < .001$, $\eta_p^2 = .26$.

To investigate the main effect of group in more detail, we again conducted one-way ANOVAs between the control group and each of the experimental groups. These comparisons revealed no significant difference between the control group and the annotation group, $F < 1$, nor between the control group and the VN group, $F(1,41) = 2.30$,

p = .137, η_p^2 = .05. However, we found a significant difference between the control group and the RN group, $F(1,43)$ = 6.26, p = .016, η^2 = .13, d = .76, the control group and the RP group, $F(1,40)$ = 17.32, p < .001, η^2 = .30, d = 1.32, and the control group and the RPF group, $F(1,44)$ = 19.01, p < .001, η^2 = .30, d = 1.31.

Note Content. Our next set of analyses focused on the content of the notes in the VN, RN, RP, and RPF groups, specifically with respect to the amount written, the presence of key points in the notes, and the effect of the presence of key points on test accuracy. The control group, who took no notes, was not included in these analyses.

Three content variables were computed. First, we counted the number of words written in the notes (the *word count* variable). Second, we computed the proportion of key points (out of 30) correctly produced within the notes (the *key-points* variable). In the case of the RP and RPF groups, who were specifically instructed to write down the key points, participants received 1 point for each key point. For the VN and RN groups, who were not specifically instructed to identify key points, we searched participants' notes for key points, again assigning 1 point to each. Duplications were not counted in any case. Finally, we investigated the potential relationship between producing key points in the notes and the likelihood of correctly answering test questions that were associated with those key points (the *correspondence* variable; see Coria & Higham, 2018a). Specifically, the correspondence variable was the conditional likelihood that a test question would be answered correctly given that the key point it tested was produced in the notes. For example, suppose a participant produced 15 (out of 30) key points in her notes, and of the 15 test questions associated with those key points, 12 were answered correctly. The correspondence score for that participant would be 0.80 (12/15).

Word count. The mean number of words included in the notes in the Annotation, VN, RN, RP, and RPF groups is shown in Figure 3. A one-way ANOVA on the number of words written by the groups showed a significant main effect, $F(4,102) = 80.68$, $p < .001$, $\eta_p^2 = .76$. As expected, the annotation group wrote significantly fewer words than all other groups, smallest $F(1,42) = 97.33$, $p < .001$, $\eta^2 = .70$, $d = 3.04$. One-way ANOVAs to breakdown the main effect revealed significant differences between the VN group and the RN group, $F(1,40) = 24.95$, $p < .001$, $\eta^2 = .38$, $d = 1.58$, the VN group and the RP group, $F(1,38) = 75.87$, $p < .001$, $\eta^2 = .67$, $d = 2.83$, and the VN group and the RPF group, $F(1,39) = 140.48$, $p < .001$, $\eta^2 = .78$, $d = 3.80$. We also found significant differences between the RN and RP groups, $F(1,42) = 9.80$, $p = .003$, $\eta^2 = .19$, $d = .97$, and the RN and RPF groups, $F(1,43) = 21.59$, $p < .001$, $\eta^2 = .33$, $d = 1.42$, but no significant differences between the RP and RPF groups, $F(1,43) = 1.24$, $p = .273$, $\eta^2 = .03$. In short, the VN group wrote more than all other groups, the RN group wrote an intermediate amount, the RP and RPF groups wrote somewhat less than the RN group, and the Annotation group, by far, wrote the least.

Key points. Figure 4 displays the mean proportion of lecture key points (total = 30) that were written in the notes as a function of group. Neither the control group, nor the annotation group, who were shown key points in their LSHs and took limited notes, were included in this analysis. A one-way ANOVA on the proportion of key points produced by the VN, RN, RP, and RPF groups revealed a main effect, $F(3,81) = 10.47$, $p < .001$, $\eta_p^2 = .28$. As shown in Figure 4, participants in the VN group correctly noted a significantly higher proportion of key points than the RN group, $F(1,40) = 16.98$, $p < .001$, $\eta^2 = .30$, $d = 1.30$, the RP group, $F(1,38) = 33.98$, $p < .001$, $\eta^2 = .47$, $d = 1.89$, and the RPF group, $F(1,39) = 37.16$, $p < .001$, $\eta^2 = .49$, $d = 1.95$. No other differences were significant, largest $F < 1$.

Thus, the tendency for participants in the VN group to write more words had the effect of increasing the proportion of key points produced in their notes compared to the other three groups. This was unsurprising given that the VN group was copying information verbatim from the slides, which contained the key points. However, the next analysis (correspondence) addressed the question of whether the presence of a high proportion of key points in the VN group's notes facilitated performance on the test questions associated with those points.

***Correspondence*.** The correspondence variable was broken down by immediate and delayed testing in order to determine any effects of forgetting over the weeklong delay.

Immediate Testing. Figure 5 shows correspondence as a function of group in the immediate testing group. A one-way ANOVA on correspondence in the VN, RN, RP, and RPF groups revealed a main effect of group, $F(3, 81) = 4.14$, $p = .009$, $\eta_p^2 = .13$. A breakdown of the main effect revealed that there was no significant difference in correspondence between the VN and RN groups, $F < 1$, but there were significant differences between the VN and the RP group, and between the VN and RPF groups, $F(1,38) = 10.57$, $p = .002$, $\eta^2 = .22$, $d = 1.05$ and $F(1,39) = 4.87$, $p = .033$, $\eta^2 = .11$, $d = .71$ respectively. There were also significant differences between the RN and RP groups and between the RN and RPF groups, $F(1,42) = 7.35$, $p = .010$, $\eta^2 = .15$, $d = .84$ and $F(1,43) = 4.08$, $p = .050$, $\eta^2 = .09$, $d = .62$, respectively. There was no difference between the RP and RPF groups, $F < 1$. Thus, even though participants in the VN group, and to a lesser extent the RN group, wrote more notes and were more likely to produce key points in their notes compared to the RP and RPF groups, the benefit of doing so was less. For example, if a key point was produced in the VN group's notes, participants were 86%

likely to answer the associated test question correctly compared to a greater than 93% likelihood in the RP and RPF groups.

Delayed Testing. Figure 5 also shows mean correspondence for each group at delayed testing. A one-way ANOVA on these data revealed a significant main effect, $F(3,76) = 10.44$, $p < .001$, $\eta_p^2 = .29$. At delayed testing, there was a significant difference between correspondence in the VN and RN groups, $F(1,38) = 11.88$, $p = .001$, $\eta^2 = .24$, $d = 1.12$, the VN and RP groups, $F(1,35) = 27.93$, $p < .001$, $\eta^2 = .44$, $d = 1.79$, and the VN and RPF groups, $F(1,39) = 24.83$, $p < .001$, $\eta^2 = .39$, $d = 1.60$. There were no other significant differences between the groups, largest $F(1,37) = 3.19$, $p = .082$, $\eta^2 = .08$. Thus, just as with immediate testing, participants in the RP and RPF groups showed the highest correspondence. However, whereas at immediate testing both the VN and RN groups showed similar correspondence scores, the VN group show considerably lower rates of correspondence at delayed testing.

Summary for Experiment 1. The results for test accuracy show that across both immediate and delayed testing, both the RP and RPF groups outperformed all other groups. Thus, of all of the methods for notetaking investigated within this research, notetaking in the form of retrieval practice during lecture pauses led to the highest test scores. Furthermore, although the VN group showed significantly higher test scores than controls at immediate testing whereas the RN group did not, this pattern reversed following a delay. However, as the immediate and delayed tests were analysed separately, it is unknown whether this interaction between the VN and RN group over time is significant. Finally, the annotation group did not perform better than the control group at either immediate or delayed testing. The note-content analyses revealed that the VN group wrote significantly more than any other group and identified the greatest number of key points in their notes. However, the VN group also showed the greatest amount of forgetting for key points between the lecture and delayed testing when compared to RN, RP and RPF groups, a finding that is consistent with the accuracy data.

A potential criticism of these findings is that the exposure duration to the lecture material varied between the groups. That is, a side effect of incorporating multiple lecture pauses for notetaking in the retrieval practice groups was that participants had longer to learn the lecture material (see Figure 1). Conceivably, this increased exposure to the material may have been the cause of the retrieval practice groups' test advantage relative to the other groups, rather than retrieval practice per se. A second concern is that the multiple exploratory measures we obtained in in Experiment 1 (retrospective confidence, lecture enjoyment, aJOL judgments, and TUTs) may have affected the accuracy results (e.g., Mitchum, Kelley, & Fox, 2016). Both of these issues are addressed in Experiment 2.

Experiment 2

To determine whether long exposure to the lecture material contributed to the retrieval practice groups' excellent performance in Experiment 1, we conducted a second experiment. This experiment was designed to fully control exposure time and included only a control group and a RPF group. We did not consider it necessary to include the any other groups beyond a control group and the RPF group, given that it was specifically the RPF group that performed well and had the greatest additional exposure to the lecture content during encoding. Hence, Experiment 2 was specifically designed to determine whether the RPF group's excellent performance would replicate under conditions where exposure time to the lecture material was matched to control. The lecture, test questions and location of the experiment were the same as those used in Experiment 1. To control exposure time, pauses were incorporated into the lectures for *both* the control and RPF groups. The control group was instructed to spend the lecture pauses thinking about the material that had just been covered in the lecture, whilst viewing a feedback slide (containing key points) that was displayed for the full duration of the pause. Participants in the RPF group, as in Experiment 1, were first instructed to write three key points from the previous lecture segment, and then were shown the key points as feedback. Critically, the total time spent retrieving and/or viewing the key points – i.e., the time that participants were exposed to the lecture material – was equated between the groups (see Method for details).

The second reason for conducting Experiment 2 was to eliminate any potential influence that eliciting multiple exploratory ratings may have had on test performance. Some research has shown that requiring participants to make metacognitive judgments can alter memory performance (e.g., Mitchum et al., 2016). Thus, by eliminating the exploratory ratings, we would also remove this potential confounding. Specifically, participants in Experiment 2 were not required to rate retrospective confidence or lecture enjoyment. Nor were they required to make aJOL judgments or indicate whether they had experienced TUTs. Instead, the procedure was simplified, and participants were free to concentrate on test performance and nothing else.

Method

Participants. In total, 38 undergraduate students from the University of Southampton (5 males, 33 females, *M* age = 19.04, *SD* age = 1.30) took part in this research in exchange for course credit. The participants were assigned to one of two encoding groups (explained in more detail later) based on the session that they signed up to attend: control = 20, RPF = 18. In Phase 2, two of the participants who were present for

Phase 1 did not return, leaving the following sample sizes for Phase 2: control = 19, RPF = 17.

Design and materials. All aspects of the design and materials were the same as Experiment 1, except for the following changes. Only two encoding groups were included in Experiment 2: the control group and the RPF group. Whilst the RPF group followed the same procedure as Experiment 1, the control group differed in that they were provided a 90-sec pause to "think about the material," and view a feedback slide (see Procedure section for more information). All other encoding groups from Experiment 1 were dropped from this experiment. All measures, except for test accuracy on the immediate and delayed tests, were also eliminated from Experiment 2. The materials provided to the control and RPF groups were the same as for the respective groups in Experiment 1.

Procedure. The procedure for the RPF group was similar to Experiment 1 except for the elimination of rating regarding TUTs, aJOLS, lecture-enjoyment, and retrospective confidence. The lecture sections for the control group were interspersed with pauses of the same length (90-sec) as the RPF group, but instead of retrieving the key points for 60-sec and viewing the feedback for 30-sec, the control group were instructed to pause and "think about the material" covered in the previous lecture section for 90-sec. During this 90-sec pause, a slide showing the three key points (identical to the feedback slide shown in the RPF group) was on the screen for the entire 90-sec duration of the pause. The control group did not take any notes and were not provided with any handout material. As in Experiment 1, the RPF group viewed each lecture section before being instructed to take 60-sec to summarize the three key points on the provided handout sheet, identical to the one provided in Experiment 1. They were then presented with the feedback slide for 30-sec.

Ethics Approval. The study was reviewed and approved by the University of Southampton Ethics Committee and the Research Governance Office. All participants were consenting adults over 18 years old and all gave informed consent for their data to be included in the study and any publications that follow from it. Following completion of the study, participants were fully debriefed.

Results and Discussion

Immediate test accuracy. The top panel of Table 4 shows mean accuracy (and 95% confidence intervals) on the immediate test as a function of encoding group and test type. As in Experiment 1, the data are divided in Table 4 according to test type (multiple choice vs cued recall) for completeness but were collapsed over test type for purpose of analysis. The overall group means, which are weighted 2:1 according to 10 multiple-choice questions and 5 cued-recall questions, are shown in the far-right column of Table 4. A one-way independent-measures ANOVA with two levels (controls vs RPF) on the collapsed immediate test-accuracy data revealed no effect of group, $F(1,37) < 1$.

Delayed test accuracy. The bottom panel of Table 4 shows mean accuracy (and 95% confidence intervals) for the delayed test as a function of encoding group and test type. As before, the analysis was conducted on the total scores. A one-way independent-measures ANOVA with two levels (group: matched controls vs matched RPF) on the total delayed test-accuracy data revealed a main effect of group, $F(1,35) = 12.66$, $p = .001$, $\eta^2 = .27$, $d = 1.20$.

In summary, Experiment 2 has shown that there were no differences between the control group and the RPF group at immediate testing, suggesting that the added time to study (rather than retrieve) the material in lecture pauses was effective in the short term.

However, much like the benefits of verbatim notetaking observed in Experiment 1, these benefits were short lived; once testing was delayed, the RPF group performed considerably better than the control group.

General Discussion

The study has produced five main findings. First, annotating an LSH provided no learning benefit over passive observation. This result is potentially surprising, given the popularity of LSHs in modern lectures. However, these findings are consistent with Coria and Higham (2018a) who also found no memorial advantages of LSH annotation compared with passive observation. Together, these results suggest that LSHs do not promote "efficient encoding" (Marsh & Sink, 2010, p. 692) and that students should be encouraged to engage in more effective learning activities during lectures.

Second, relative to passive controls, verbatim notetaking led to better immediate test performance, suggesting that something akin to the production effect (MacLeod et al., 2010) may have been occurring in the short term. However, this advantage disappeared after a delay. Although this has not been confirmed with a statistical test simultaneously comparing immediate and delayed performance for both verbatim and regular notetaking, these results appear to replicate analogous findings obtained by Bui et al., (2013). In a similar vein, Mueller and Oppenheimer (2014) found that taking notes on a laptop led to more verbatim transcription, which, in turn, led to poorer test performance than those who took longhand notes. Their retention interval was only approximately 30-min, which suggests that the advantage of verbatim notetaking can be quite short lived or that short-term facilitation from verbatim notetaking does not occur in all circumstances.

The content analyses of our participants' notes revealed that verbatim note takers wrote substantially more than any other group and identified a considerably greater proportion of key points in their notes. Identifying these points appeared to assist participants in the short term; participants in the VN group answered correctly most of the test questions associated with those key points (i.e., correspondence = .86). However, by delayed testing, having identified those key points earlier was not of much help. That is, correspondence decreased dramatically over the delay, from .86 to .57, the largest drop of any of the groups we tested (cf. Figure 5). This result suggests that the act of reproducing the important lecture information in notes is not sufficient for it to be retained in memory and/or to be used to answer corresponding test questions, at least after a delay.

Our third major result was that, with respect to the effect of delay, regular notetaking appeared to show the opposite pattern to verbatim notetaking. That is, regular note takers did not perform significantly better than passive controls at immediate testing (although there was a trend in that direction), but they did show significantly better performance after a delay. This result implies that regular notetaking leads to longer-term retention of key points than verbatim notetaking. This conclusion is supported by the correspondence results from the notetaking analysis; whereas correspondence in the VN group decreased by .29 over the delay (.86 - .57), it only decreased by approximately half that amount (.15) in the RN group (.86 - .71).

Our fourth finding was that providing retrieval-practice opportunities in lecture pauses produced the highest test scores at both immediate and delayed testing. This result indicates that incorporating a retrieval component to notetaking is the most effective strategy of those included in this research. Such results are not unprecedented.

Several studies have shown that taking several short quizzes throughout a learning session (interpolated testing) can produce both retrospective memorial benefits (i.e., good memory for the material that is tested) and prospective learning benefits (i.e., better learning of new material following interpolated testing; e.g., Szpunar, McDermott, & Roediger, 2008; Jing, Szpunar & Schacter, 2016) Regarding the former benefit, as noted earlier, Szpunar et al. (2013) found that interpolated testing led to superior final, cumulative test performance compared to interpolated restudying. Our research, then, can be considered further support for this finding in the context of notetaking.

Finally, providing feedback in the RPF group did not further enhance test scores compared to the RP group. This result was unexpected given the literature demonstrating a benefit of corrective feedback (e.g., Butler & Roediger, 2007; Griffiths & Higham, 2018; Kornell et al., 2009; McDaniel & Fisher, 1991; Potts & Shanks, 2014; Yang, Potts, & Shanks, 2017). One potential explanation for the null effect of feedback is that there were few errors made by participants. Over 70% of key points were correctly identified in the notes of the retrieval-practice groups and feedback can be largely superfluous if performance is good (Butler & Roediger, 2007; Kang, McDermott & Roediger, 2007; Pashler, Cepeda, Wixted, & Rohrer, 2005). However, one might reason that there is still some room for improvement at that level of performance, so it is not clear that high performance is the sole cause that feedback had little effect. Another potential factor was the length of presentation of the feedback (30 seconds). Potentially, if participants were already having some difficulty keeping up with the lecture and engaging in repeated retrieval-practice attempts, processing three key points over a 30 second interval could have been overwhelming and not sufficient time for participants to process the information (see Butler & Roediger, 2007). However, this explanation also falls somewhat short in that being provided with 10 seconds to process each key point might be considered enough time. Future research should investigate further the effect of feedback on retrieval-practice attempts to provide a better answer to this quandary.

A potential criticism for the excellent test performance observed in both retrieval-practice groups in Experiment 1 is that the overall length of the lecture had been increased to incorporate the necessary pauses. Although research such as Carrell, Dunkel and Mollaun (2004) indicates that learning material over a shorter time can actually improve recall, particularly when taking notes, it could be postulated that the effect of increased time spent learning was the cause of the better performance in the RP and RPF groups compared to the other groups. However, in Experiment 2, the variations in timing were removed and the pattern of better performance for the group engaging in retrieval practice versus passive observation persisted following delay. In other words, when students engaged in passive observation but were given pauses (of equal length to those engaging in retrieval practice and receiving feedback) and indeed were encouraged to spend time thinking about the lecture content, their test performance following delay was less compared to participants who engaged in retrieval practice and received feedback. These results suggest that any effect found in Experiment 1 due to increased exposure to lecture material in the RP and RPF groups was short lived, and that the cause of the improved long-term test performance as measured by the delayed test, was the encoding method and not prolonged exposure to the lecture material. It is also worth noting that the delayed test scores for both controls and RPF groups across both experiments were remarkably similar (cf. relevant conditions in Tables 3 and 4). Although descriptive, these results also indicate internal consistency in the results.

Theoretical Account of the Results

We believe that the main results from our research can largely be explained by Bjork and Bjork's (1992) *New Theory of Disuse* and the related principle of desirable difficulty (Bjork, 1994). The new theory of disuse is a modification of Thorndike's (1914) law of disuse, which proposes that memory traces decay over time if unused, and incorporates interference with other information, and the effects of disuse. Within this theory, there are two types of memory strength: retrieval strength and storage strength. Retrieval strength refers to the current accessibility of information in memory. It varies depending on the context of retrieval and reflects how primed a memory representation is due to recency and the retrieval cues that are present at the time of retrieval. The level of retrieval strength completely determines whether an item will be recalled or not. Storage strength, on the other hand, is an index of learning but has no direct effect on memory performance. It reflects how well an item is interrelated with other items in memory and increases with repeated study and retrieval opportunities. Whereas retrieval strength can increase or decrease depending on the retrieval context and other factors, the storage strength can only increase, not decrease, once it is accumulated.

Importantly, although storage and retrieval strength reflect separate components of memory, they interact with each other when items are remembered. For example, storage strength increases as a monotonic function of study and retrieval opportunities, with the latter generally increasing storage strength more than the former. However, the increment to storage strength that is caused by retrieval is dependent on retrieval strength. Specifically, storage strength is increased more if retrieval strength is low rather than high, so the more difficult retrieval is, the more beneficial it will be when the item is remembered. In other words, the amount of learning (increase to storage strength) caused by retrieval practice will be greater if the retrieved item is less accessible. This underpins one aspect of the second principle - desirable difficulty: difficult retrieval confers more benefits to learning compared to facile retrieval.

Retrieving or studying an item does not just increment storage strength but increments retrieval strength as well in the short term; an item that has just been studied or retrieved is more accessible afterwards. Again, it is assumed that retrieving information has a greater incremental effect than studying it. However, unlike storage strength, this increased retrieval strength can be lost. When trying to retrieve this same information following interference or a time delay, the retrieval strength can again be low, thus storage strength will be further increased. This fact explains how information that has exceptionally high retrieval strength at one time (e.g., one's telephone number during childhood), can be hard or impossible to retrieve at a later time (i.e., in adulthood, after a period of disuse) and following delay can be difficult to retrieve again. Critically, the amount that retrieval strength reduces due to disuse (or learning new items) is impacted by both the current level of retrieval strength and storage strength. In particular, as current levels of retrieval strength increase, disuse causes *greater* decreases in retrieval strength. Conversely, as current levels of storage strength increase, disuse causes *lesser* decreases in retrieval strength.

In terms of our results at immediate testing, we need only assume that verbatim notetaking (in the VN group) and retrieval practice (in the RP and RPF groups) both increase retrieval strength by about the same amount, leading to good immediate recall, shown by test scores significantly higher than the control groups. Although one might expect retrieval practice to increase retrieval strength more than copying information, retrieval strength reflects current memory accessibility. It therefore seems quite plausible that having just written down (i.e., "produced"; MacLeod et al., 2010) most of the key points would render those key points accessible in the short term, which would facilitate immediate test performance. Regular notetaking also increments retrieval strength, but not as much as in the VN, RP or RPF groups. Presumably, the increment to retrieval strength is not as great because the RN group is similar to the RP and RPF groups, but with less desirable difficulty. Specifically, whereas the RP and RPF groups were required to wait until lecture pauses before taking notes, the RN group could rely more on working memory to take notes, only occasionally needing to retrieve information from long-term memory to complete their notes. The delays before taking notes would have meant that retrieval was more difficult in the RP/RPF groups vs the RN group, leading to greater increments in retrieval strength, and better immediate test performance. Given the equivalence of the control and annotation groups on the immediate test, annotating LSHs failed to increment retrieval strength any more than simply listening to the lecture.

At delayed testing, however, the storage strength must be considered. A reasonable assumption of Bjork and Bjork's (1992) theory is that storage strength is increased more by difficult retrieval (in the RP, the RPF, and to a lesser extent, the RN group) than by verbatim copying (VN group). Difficult retrieval would also increase storage strength more than LSH annotation (annotation group) or passively listening (control). High levels of storage strength protect against forgetting. It is therefore telling that the RN, RP and RPF forgot the least (immediate test performance minus delayed test performance = .17 for all three groups), the control and annotation groups forgot somewhat more (.20–.22) whereas the VN group forgot the most (.26). The reason why the VN group forgot more than the annotation and control groups despite all having low storage strength was because the VN group had particularly high retrieval strength at immediate test, whereas the other two groups did not, which meant they had “more to lose” over the delay. Thus, the New Theory of Disuse appears to provide a full account of the test performance we observed at both immediate and delayed testing. Bui et al. (2013) also referred to this theory to account for their verbatim vs regular notetaking results.

Recommendations for Educators

The results of our research allow us to make a number of recommendations for educators. First, it may be necessary to reconsider the common practice of giving students access to LSHs *prior* to lectures. Many students will bring LSHs to lectures to help them encode the lecture material, potentially adding a few annotations to the printed or digital copies in their possession. The current research, coupled with previous findings (e.g., Coria & Higham, 2018a; Marsh & Sink, 2010), suggests that students take far fewer notes when LSHs are available. To the extent that LSH annotation replaces notetaking, encoding of lecture material may be undermined.

However, provision of LSHs *following* lectures may still be good practice as long as students continue to take notes in lectures, particularly if those notes are part of retrieval-practice opportunities. LSHs provide a good external storage record of lecture material and may be helpful in revising for exams, so there is no need to withhold them from students as long as their presence does not undermine notetaking behavior.

Verbatim notes also provide a good external storage record as they are replicas of the LSHs. In this vein, Bui et al. (2013, Experiment 3) compared test performance between groups analogous to our RN and VN groups, only they manipulated whether participants were allowed to review the notes they had taken prior to taking a test 24 hours later. They found that if participants were not allowed to review the notes, regular notetaking was better than verbatim notetaking, a result similar to their earlier experiment and our current results. However, if participants were allowed to review their notes, the pattern was reversed. Presumably this latter outcome occurred because verbatim notes provided a more complete external storage record compared to more selective regular notes.

However, in our view, this result is somewhat of a red herring in that it is only in rare cases in modern educational settings that students rely solely on their own notes when revising for exams. They will usually also have access to a textbook or other readings, LSHs, and possibly other revision material prepared by the lecturer that would be the same for all students regardless of their notetaking strategy. In terms of the encoding function, which was our primary focus in this research, it is clear that students should be dissuaded from taking verbatim notes. After a delay, this type of notetaking is no better for learning than passively observing the lecture (or annotating LSHs).

The real question to us is whether there are differential *savings* between different notetaking styles when it comes to revision. For example, would students who take verbatim notes need longer to revise compared to those who have taken notes in a way that incorporates retrieval practice? If so, there may be latent advantages to encoding lecture material in superior ways despite the fact that all students have access to the same materials during revision.

Coria and Higham (2018b) investigated this possibility in recent study. Participants first encoded a lecture with either LSH annotation, passive observation, or regular notetaking. After an eight-week long delay, all students returned to the lab for a memory test. In one experiment, participant simply wrote a final test with no revision. The results in this group were similar to those obtained here and in Coria and Higham (2018a): the notetaking group outperformed both the control and annotation groups, which did not differ. These results suggest that the encoding advantages of notetaking persevere over the long term. However, in another experiment, students were permitted to revise for the test for as long as they liked by studying an unmarked copy of the LSH. The LSH contained all the key information from the lecture that was to be tested. In contrast to the first experiment, there was no difference in test performance between the three groups; the opportunity to revise appeared to nullify any advantages attributable to notetaking. However, the story was quite different if revision time was examined. Specifically, students who took regular notes spent considerably less time preparing for the test compared to those who annotated LSHs or controls, who did not differ in their revision time. Thus, even after an eight-week retention interval and a revision opportunity, the encoding advantages of notetaking can still be observed, albeit in a subtle form.

Although regular notetaking conferred advantages relative to LSH annotation and verbatim notetaking (after a delay), the groups that practiced retrieving key points during lecture pauses (RP and RPF groups) outperformed all other groups at both immediate and delayed testing. Consequently, we strongly recommend that educators adopt an activity analogous to this in their lectures. One potential criticism of this activity is that it interrupts the flow of the lecture and takes up valuable time that could be used to present additional material. One possible solution to this criticism is to make retrieval practice part of students' homework assignment, to be completed outside of lectures rather than during lectures.

On balance, we would recommend incorporating notetaking into lectures that takes the form of multiple retrieval-practice opportunities as a method of improving retention of information. In our study, we found a limited benefit of providing feedback following retrieval practice. Thus, providing the extra time for participants to view feedback was not beneficial within our research. However, according to the bifurcation distribution model (Kornell et al., 2011), retrieval practice without feedback causes retrieved items to become stronger but non-retrieved items to remain weak. To ensure that non-retrieved items also gain some strength during retrieval-practice opportunities, it may be necessary to include feedback. Feedback would also ensure that errors of commission are not strengthened should they occur. Such errors are likely amongst low-performing students, so for these students, feedback may be essential. We therefore, despite our findings, recommend incorporating feedback following the retrieval-practice opportunities.

Conclusion

Our study builds on the previous work of Coria and Higham (2018a), which demonstrated that longhand notetaking promotes memory of lecture material more than lecture-slide handout annotation. Furthermore, we have been able to identify an additional strategy that increases the beneficial effects of notetaking that is simple to implement in real lectures: guide students to avoid taking notes immediately, but rather to wait and practice retrieval of lecture material when given the opportunity. Such a strategy confers both memorial and metacognitive encoding benefits. Overall, we believe that educators should take a more active role in designing lectures to encourage good notetaking practices that incorporate retrieval practice. Whilst lecturers cannot control how much students choose to revise on their own, they can provide both a desirably difficult and efficient encoding experience that can, in turn, promote improved memory in later testing.

Paper 2 - Tables

Table 1

The Materials and Instructions Given to the Six Encoding Groups in Experiment 1 at the Beginning of the Experiment (All Groups) and Following Retrieval Practice (RPF Group Only)

Group	Materials	Instructions
Control	None	“Please watch this lecture”
Annotation	A pen and replicas of the lecture slides on A4 paper (six pages with six slides per page laid out in a 3 x 2 format, except for the final page with two slides; 32 slides in total)	“Please watch this lecture and annotate the handout as you would in a typical lecture”
Verbatim Notes (VN)	A pen and blank A4 paper	“Please watch this lecture and use your pen and paper to write down as much of the slide content as possible”
Regular Notes (RN)	A pen and blank A4 paper	“Please watch this lecture and use your pen and paper to take notes as you would in a normal lecture”

Retrieval Practice (RP)	A pen and an A4 printout of 10 boxes (one for each section of the lecture), each with the numbers 1-3 left-justified in each box for the participants to write three key points.	“Please watch this lecture. Do not write until the voiceover instructs you to do so at the end of each section. Then, please write the three key points – the points you think are the most important from the previous section - in the allocated pause”
Retrieval Practice with Feedback (RPF)	As in RP Group	*Prior to the lecture:* As in RP Group *When feedback was shown following retrieval practice:* “These are the three key points that you should have written. Please take 30 seconds to read them and see if they are the same as your answers”

Table 2

Mean (Standard Deviation) [±95% Confidence Limits] Accuracy on the Immediate Memory Test in Experiment 1 as a Function of Group and Test Type

Group	Test Type		
	Multiple Choice	Cued Recall	Total
Control	.81 *(.17)* [.76, .85]	.54 *(.25)* [.46, .62]	.72 *(.20)* [.66, .78]
Annotation	.79 *(.13)* [.74, .85]	.58 *(.26)* [.48, .69]	.72 *(.15)* [.65, .79]
VN	.91 *(.08)* [.85, .96]	.70 *(.20)* [.58, .81]	.84 *(.09)* [.76, .90]
RN	.89 *(.12)* [.84, .95]	.57 *(.26)* [.47, .67]	.79 *(.14)* [.71, .85]
RP	.91 *(.09)* [.86, .97]	.76 *(.21)* [.66, .87]	.86 *(.10)* [.79, .93]
RPF	.93 *(.11)* [.88, .99]	.72 *(.24)* [.62, .82]	.86 *(.11)* [.80, .93]
Total	.87 *(14)* [.85, .90]	.65*(.25)* [.60, .69]	

Note: The means in "Total" column are based on 10 multiple-choice and 5 cued-recall questions. VN = verbatim notes, RN = regular notes, RP = retrieval practice, RPF = retrieval practice with feedback.

Table 3

Mean [±95% Confidence Limits] Accuracy on the Delayed Memory Test in Experiment 1 as a Function of Group and Test Type

	Test Type		
Group	Multiple Choice	Cued Recall	Total
Control	.62 *(.18)*	.32 *(.16)*	.52 *(.14)*
	[.56, .68]	[.23, .40]	[.46, .58]
Annotation	.62 *(.16)*	.26 *(.24)*	.50 *(.15)*
	[.55, .68]	[.17, .35]	[.43, .56]
VN	.67 *(14)*	.34 *(.16)*	.58 *(.11)*
	[.61, .74]	[.25, .44]	[.52, .63]
RN	.76 *(.13)*	.37 *(.25)*	.62 *(.13)*
	[.70, .82]	[.27, .46]	[.56, .68]
RP	.79 *(.15)*	.48 *(.18)*	.69 *(.12)*
	[.73, .86]	[.38, .58]	[.63, .75]
RPF	.79 *(.12)*	.48 *(.24)*	.69 *(.13)*
	[.73, .85]	[.39, .57]	[.63, .75]
Overall	.70 *(.16)*	.38 *(.22)*	
	[.68, .73]	[.34, .41]	

Note: The means in "Total" column are based on 10 multiple-choice and 5 cued-recall questions. VN = verbatim notes, RN = regular notes, RP = retrieval practice, RPF = retrieval practice with feedback.

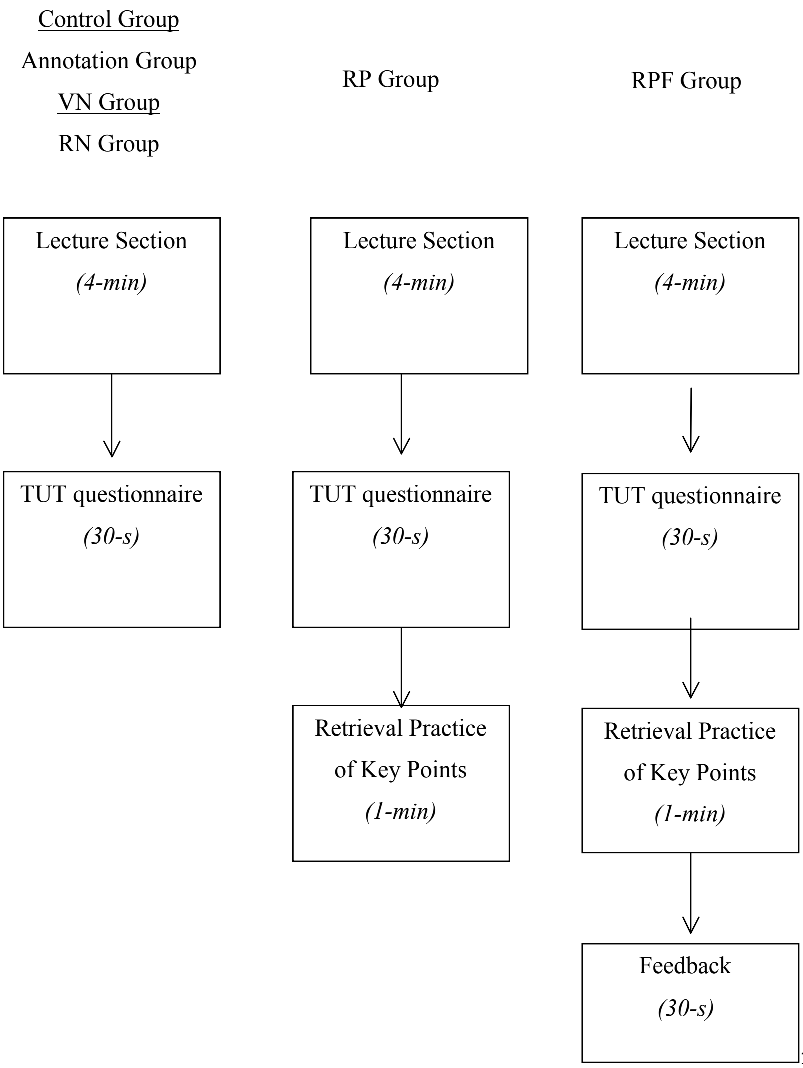
Table 4

Mean (Standard Deviation) [±95% Confidence Limits] Accuracy on the Test in Experiment 2 as a Function of Group, Delay, and Test Type

	Test Type		
Group	Multiple Choice	Cued Recall	Total
		Immediate Test	
Control	.91 *(.11)*	.57 *(.57)*	.79 *(.11)*
	[.86, .95]	[.47, .67]	[.74, .84]
RPF	.87 *(.10)*	.68 *(.68)*	.81 *(.11)*
	[.82, .92]	[.58, 79]	[.76, .86]
		Delayed Test	
Control	.66 *(.19)*	.23 *(.20)*	.52 *(.17)*
	[.58, .74]	[.12, .34]	[.44, .59]
RPF	.79 *(.15)*	.54 *(.26)*	.71 *(.15)*
	[.70, .87]	[.43, .66]	[.63, .78]

Note: RPF = retrieval practice with feedback

Paper 2 - Figures



10 l

Figure 1. The tasks and timings involved in each of the lecture blocks in Experiment 1 as a function of encoding group. VN = verbatim notes, RN = regular notes, RP = retrieval practice, RPF = retrieval practice with feedback, TUT = task unrelated thoughts.

2) In which of the following environments would MRSA create the biggest problem?

	A	B	C	D	E
Answers:	Prison	School	Outdoors	A family house	Hospital
Rank:	2	3	5	4	1
Confidence:					95%

3) MRSA is normally asymptomatic, what happens to a person to cause the onset of symptoms?

Answer	**Confidence**
A weakened immune system	65%

Figure 2. Sample multiple-choice (top panel) and cued-recall (bottom panel) test questions. Both questions were answered correctly.

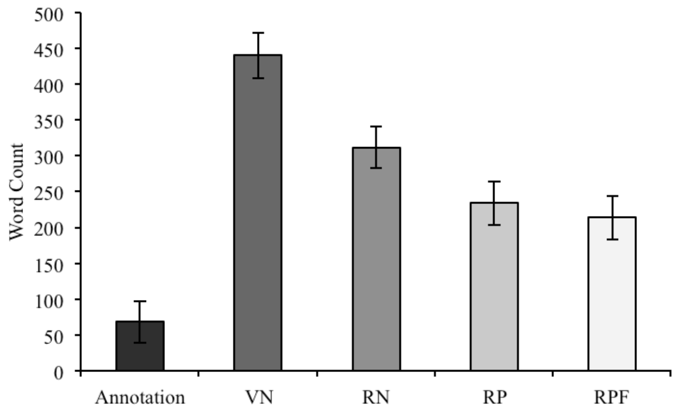


Figure 3. Mean number of words written as notes during the lecture. VN = verbatim notes, RN = regular notes, RP = retrieval practice, RPF = retrieval practice with feedback. Error bars indicate 95% confidence intervals.

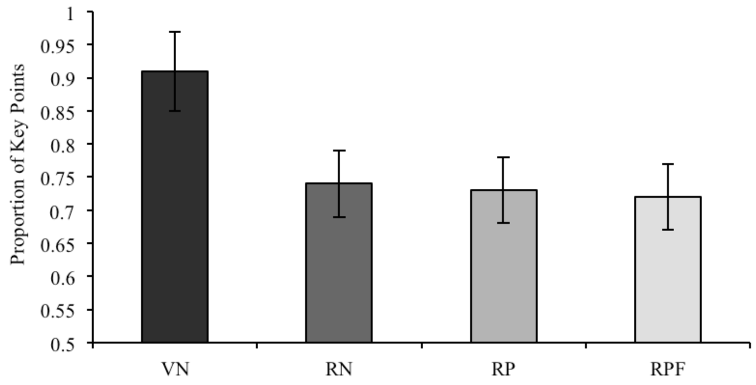


Figure 4. Mean proportion of key points produced in participants' notes as a function of group. VN = verbatim notes, RN = regular notes, RP = retrieval practice, RPF = retrieval practice with feedback. Error bars indicate 95% confidence intervals.

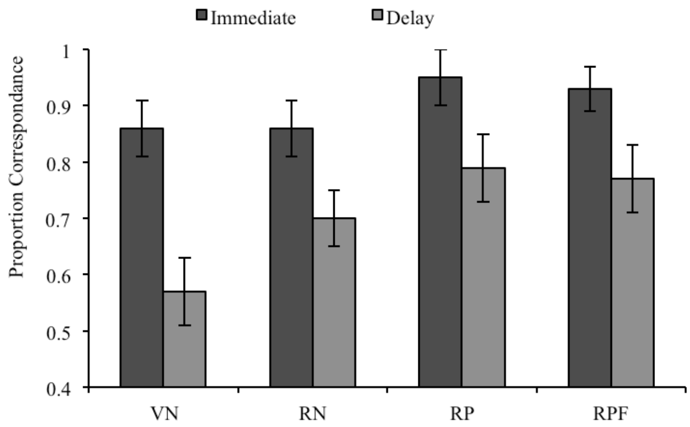


Figure 5. Correspondence at immediate and delayed testing (the conditional likelihood that an immediate test question would be answered correctly given that the key point it tested was produced in the notes). VN = verbatim notes, RN = regular notes, RP = retrieval practice, RPF = retrieval practice with feedback. Error bars indicate 95% confidence intervals.

Paper 2 – Supplementary Findings

Retrospective Confidence Ratings

Table 1 shows mean confidence ratings (and 95% confidence intervals) for the immediate-test answers as a function of encoding group and test type. The data are split according to test type (multiple choice vs cued recall) in Table 1 for completeness, but were collapsed over test type for purposes of analysis. The overall group means, which are weighted according to 10 multiple-choice questions and 5 cued-recall questions, are shown in the far-right column in Table 4. A one-way, between-subjects ANOVA with six levels (group: control, annotation, VN, RN, RP, RPF) on these collapsed data revealed a main effect of group, $F(5,135) = 5.76$, $p < .001$, $\eta_p^2 = .18$.

We conducted pairwise comparisons (as one-way between subjects ANOVAs) between the control groups and each of the experimental groups to investigate the main effect of group in more detail. These comparisons revealed no significant differences between the control group and the annotation group, $F < 1$, nor between the control group and the VN group, $F(1,51) = 2.60$, $p = .113$, $\eta^2 = .05$. However, we found a significant difference between the control group and the RN group, $F(1,55) = 5.20$, $p = .027$, $\eta^2 = .09$, $d = .62$, the control group and the RP group, $F(1,53) = 13.42$, $p = .001$, $\eta^2 = .20$, $d = 1.01$, and the control group and the RPF group, $F(1,54) = 15.55$, $p < .001$, $\eta^2 = .22$, $d = 1.07$.

Table 2 shows mean confidence ratings (and 95% confidence intervals) for the delayed-test answers as a function of encoding group and test type. A one-way, between-subjects ANOVA with six levels (group: control, annotation, VN, RN, RP, RPF) on the collapsed confidence ratings at delayed testing (far-right column in Table 5) revealed a main effect of group, $F(5,118) = 7.03$, $p < .001$, $\eta_p^2 = .23$.

One-way ANOVAs between the control group and each of the experimental groups revealed, as with the confidence ratings for immediate testing, there were no significant differences between the control group and the annotation group, $F < 1$, nor between the control group and the VN group, $F < 1$. However, we found a significant difference between the control group and the RN group, $F(1,43) = 4.58$, $p = .038$, $\eta^2 = .10$, $d = .65$, the control group and the RP group, $F(1,40) = 29.14$, $p < .001$, $\eta^2 = .42$, $d = 1.71$, and the control group and the RPF group, $F(1,44) = 12.12$, $p = .001$, $\eta^2 = .22$, $d = 1.05$.

Summary for retrospective confidence. The results with confidence at immediate and delayed testing generally mirrored accuracy. The exception was that at immediate testing, compared to controls, the VN group showed significantly higher accuracy, but not significantly higher confidence, whereas the RN group showed significantly higher confidence but not significantly higher test accuracy.

However, the overall pattern of confidence and accuracy observed here for the RP and control groups differed somewhat from the earlier research of Coria and Higham (2018). They found that accuracy was better in the RP group than the control group, but confidence was equated, suggesting that confidence ratings were not sensitive to the accuracy benefits that notetaking afforded. In contrast, in our current research, accuracy was again higher in the RP group, at least at delayed testing, but so was confidence. One potential explanation is that notetaking in longer lectures, such as the one used here, provided valid cues for confidence that participants were able to utilize that are either not available or not utilized with shorter lectures such as those used in Coria and Higham (2018). Given the importance of this issue for self-regulated learning – students will tend to engage in learning strategies that they *believe* are efficacious and which boost their confidence – future research should explore these confidence-accuracy relationships in more detail.

Aggregate Judgments of Learning

Due to experimenter error, eight participants in the annotation group were not provided with sheets on which to report aJOLs. These participants were dropped from all analyses involving aJOLs.

Mean aJOLs for each of the six groups are shown in Table 3 (left column). A 6 (group: control, annotation, VN, RN, RP, RPF) x 10 (lecture sections 1 – 10) mixed model ANOVA was conducted on the aJOLs and revealed a main effect of group, $F(5,127) = 7.76$, $p < .001$, $\eta_p^2 = .23$, a main effect of lecture section, $F(1,127) = 20.58$, $p < .001$, $\eta^2 = .13$, $d = .81$, but no interaction, $F(5,127) = 1.34$, $p = .068$, $\eta_p^2 = .05$. Because we had no specific hypothesis pertaining to how aJOLs might vary over lecture section, the focus of the remaining analyses was restricted to the main effect of group.

One-way ANOVAs to conduct pairwise comparisons between the aJOLs given by the control group and each of the experimental groups indicated that there were no significant differences between the control group and the annotation group, nor between the control group and the VN group, both *F*s < 1. However, there were significant differences between the control group and the RN group, $F(1,55) = 22.82$, $p < .001$, $\eta^2 = .29$, $d = 1.29$, the control group and the RP group, $F(1,53) = 14.95$, $p < .001$, $\eta^2 = .22$, $d = 1.06$, and the control group and the RPF group, $F(1,54) = 5.35$, $p = .025$, $\eta^2 = .09$, $d = .63$. Although they did not show the highest test accuracy, descriptively the RN group gave the highest predictions for later test performance.

Summary of aggregate judgments-of-learning. The aJOLs were in line with both retrospective confidence ratings and accuracy at delayed testing. That is, the RN, RP, and RPF groups had the highest accuracy, whereas the annotation, control groups had the lowest, and this was reflected in their prospective memory predictions. Thus, just as with retrospective confidence, participants in our current research appeared to have some metacognitive insight into the learning strategies that were most effective, even before they wrote any of the tests.

Task Unrelated Thoughts

Mean Task Unrelated Thought (TUT) ratings for each of the six groups are shown in Table 3 (right column). A 2 X 10 ANOVA on these ratings revealed a main effect of group, $F(5,135) = 6.44$, $p < .001$, $\eta_p^2 = .19$ and a main effect of lecture section, $F(1,135) = 39.67$, $p < .001$, $\eta^2 = .23$, $d = 1.08$, but no interaction, $F(5,135) = 1.08$, $p = .328$, $\eta_p^2 = .04$. As with aJOLs, because we had no specific hypothesis pertaining to how TUTs might vary over lecture section, the focus of the remaining analyses was restricted to the main effect of group.

One-way ANOVAS to conduct pairwise comparisons between the TUT ratings provided by the control group and each of the experimental groups revealed that the control group showed significantly more TUTs than the annotation group, $F(1,46) = 10.22$, $p = .002$, $\eta^2 = .16$, $d = .94$, the VN group, $F(1,51) = 17.88$, $p < .001$, $\eta^2 = .26$, $d = 1.18$, the RN group, $F(1,55) = 23.40$, $p < .001$, $\eta^2 = .30$, $d = 1.30$, the RP group, $F(1,53) = 8.36$, $p = .006$, $\eta^2 = .13$, $d = .40$, and the RPF group, $F(1,54) = 17.88$, $p = .009$, $\eta^2 = .24$, $d = 1.15$.

Summary of task-unrelated thoughts. The higher TUT ratings in the control group compared to the annotation group is the only potential benefit that LSH annotation produced in our research over passive observation. However, despite having fewer TUTs, participants' accuracy in the annotation group was no better than controls. Interestingly, the retrieval-practice groups (RP and RPF) had descriptively more TUTs than the RN, VN, and annotation groups, perhaps because they were not engaged in notetaking except during the lecture pauses. However, these two groups scored the highest on both the immediate and delayed tests, which again attests to the efficacy of notetaking in the form of retrieval practice during lecture pauses.

Paper 2 – Supplementary Findings Tables

Table 1.

Mean (Standard Deviation) [±95% Confidence Limits] Confidence on the 100-point scale for Test Answers at Immediate Testing as a Function of Test Type and Group

	Test Type		
Group	Multiple Choice	Cued Recall	Total
Control	79.10 *(23.04)*	50.28 *(25.40)*	69.50 *(19.65)*
	[74.31, 83.90]	[43.16, 57.41]	[63.70, 75.30]
Annotation	79.47 *(15.56)*	56.77 *(20.99)*	71.87 *(15.77)*
	[73.50, 85.43]	[47.92, 65.63]	[64.66, 79.08]
VN	83.97 *(15.60)*	64.84 *(25.13)*	77.60 *(16.79)*
	[77.56, 90.39]	[55.31, 74.37]	[69.52, 85.68]
RN	86.77 *(7.57)*	63.57 *(20.51)*	78.98 *(10.14)*
	[80.93, 92.60]	[54.90, 72.27]	[72.54, 85.42]
RP	89.66 *(8.86)*	76.99 *(16.34)*	85.44 *(10.65)*
	[83.55, 95.76]	[67.92, 86.06]	[78.57, 92.30]
RPF	91.33 *(7.56)*	74.23 *(17.26)*	85.60 *(8.17)*
	[85.37, 97.30]	[65.38, 83.09]	[79.22, 91.78]

Note: The means in "Total" column are based on 10 multiple-choice and 5 cued-recall questions. VN = verbatim notes, RN = regular notes, RP = retrieval practice, RPF = retrieval practice with feedback.

Table 2.

Mean [±95% Confidence Limits] Confidence on the 100-point scale for Test Answers at Delayed Testing as a Function of Test Type and Group

	Test Type		
Group	Multiple Choice	Cued Recall	Total
Control	56.63 *(18.40)*	26.18 *(16.79)*	46.48 *(15.54)*
	[50.03, 63.23]	[17.62, 34.73]	[40.31, 52.65]
Annotation	57.58 *(15.10)*	30.57 *(21.34)*	48.57 *(14.27)*
	[50.35, 64.80]	[21.19, 39.95]	[41.81, 55.33]
VN	58.70 *(16.82)*	33.28 *(15.51)*	50.23 *(14.20)*
	[51.28, 66.11]	[23.67, 42.90]	[43.29, 57.16]
RN	68.09 *(18.09)*	35.28 *(24.15)*	57.14 *(17.87)*
	[61.04, 75.15]	[26.13, 44.43]	[49.80, 64.47]
RP	80.70 *(11.43)*	45.68 *(19.60)*	69.02 *(9.77)*
	[73.08, 88.32]	[35.78, 55.55]	[62.64, 75.40]
RPF	71.06 *(16.13)*	47.62 *(26.94)*	63.24 *(17.12)*
	[64.16, 77.95]	[38.68, 56.56]	[56.23, 70.25]

Note: The means in "Total" column are based on 10 multiple-choice and 5 cued-recall questions. VN = verbatim notes, RN = regular notes, RP = retrieval practice, RPF = retrieval practice with feedback.

Table 3.

Means (Standard Deviation) [±95% Confidence Limits] for Aggregate Judgments of Learning (0-100%) and Task Unrelated Thoughts (0-10)

	Variable	
Group	aJOLs	TUTs
Control	43.97 *(16.07)*	5.28 *(1.75)*
	[38.70, 49.25]	[4.72, 5.83]
Annotation	39.36 *(14.64)*	3.75 *(1.73)*
	[31.14, 47.59]	[3.06, 4.44]
VN	48.16 *(16.48)*	3.15 *(1.75)*
	[41.10, 55.22]	[2.41, 3.90]
RN	62.93 *(12.35)*	3.17 *(1.36)*
	[56.51, 69.34]	[2.50, 3.85]
RP	61.33 *(16.35)*	3.89 *(1.68)*
	[54.62, 68.05]	[3.18, 4.60]
RPF	54.30 *(16.66)*	4.05 *(1.47)*
	[47.73, 60.86]	[3.36, 4.74]

Note: VN = verbatim notes, RN = regular notes, RP = retrieval practice, RPF = retrieval practice with feedback. aJOLs = Aggregate Judgments of Learning, TUTs = task unrelated thoughts.

Paper 3 - Overview

Following on from the findings in Paper 1 and Paper 2, that longhand notetaking during encoding leads to better performance at both immediate testing, and testing following a weeklong delay, we created two studies to investigate whether the benefits of longhand notetaking would persist in a more realistic scenario. In order to achieve this whilst continuing to use laboratory studies to allow for control of extraneous factors, we amended two aspects of the previous research.

Firstly, in Experiment 1, we increased the length of delay between the encoding period and the delayed test. Given that university lectures typically run over an entire semester before a final test, we theorized that an eight-week delay would be more typical between learning and testing. Research into other beneficial learning strategies such as retrieval practice (Butler & Roediger, 2007) and successive relearning (Rawson & Dunlosky, 2013) have persisted over a longer delay than the typical 7-day delay seen in cognitive psychology experiments, so we predicted that longhand notetakers would continue to outperform slide annotators and controls. This would provide further evidence for the applied benefit of longhand notetaking within the lecture theatre.

Secondly, in Experiment 2, we incorporated a restudy phase into the experiment prior to testing. Given that research into other beneficial learning strategies, such as retrieval practice, are attenuated by restudy (see Storm, Friedman, Muryama & Bjork, 2014), we predicted that the benefits associated with longhand notetaking at the point of encoding would be mitigated by restudy. However, we also timed how long participants spent in restudy, as we predicted that the increased learning from the encoding would result in a savings effect, and participants in the longhand notetaking group would need less time engaging in restudy than the slide annotation or control groups. If this is found to be true, this is another important benefit of longhand notetaking which would be of interest to educators.

Paper 3 – The Long-term Benefits of Longhand Notetaking

Manuscript prepared for journal submission

Abstract

Coria and Higham (2018a) and Coria and Higham (2018b) demonstrated the efficacy of longhand notetaking as an encoding strategy compared to lecture slide handout annotation and passive observation. In this study, we replicated those findings over an extended retention interval between encoding and testing (Experiments 1 and 2), and with an added timed restudy phase between learning and testing (Experiment 2). We amended the design in this way to allow for a more naturalistic environment that could be generalized more easily to a university setting.

Results showed that longhand notetakers continued to outperform slide annotators and passive observers, despite the longer delay between encoding and testing, but following restudy, performance across the three groups leveled out. However, longhand notetakers needed less time in restudy than the other groups, demonstrating a further benefit of the strategy.

Keywords: *Lecture Activities, Encoding, Notetaking, Restudy*

The Long-term Benefits of Longhand Notetaking

The format of lecture delivery in universities has changed considerably as technology has developed over the past 25 years. Alongside the move from projectors and chalkboards to using software such as Microsoft PowerPoint, the availability of computers and the internet, as well as student demand (Douglas, Douglas & Barnes, 1996) has meant that it is considered good practice for lecture content to be made available to students prior to the lecture. This content is often in the form of lecture slides in either digital or paper form that can be annotated during lectures. Throughout this paper, we will refer to this as a slide handout.

Whilst these slide handouts are likely neater, more complete and potentially more accurate than the students' handwritten notes (Chen, 2013), their presence during lectures (i.e., at the point of encoding the lecture information) affects the activities that students engage in. Specifically, its presence changes the way in which students make notes (e.g., Marsh & Sink, 2010). Coria and Higham (2018a) demonstrated that providing students with a slide handout significantly reduces the amount that they engage in longhand notetaking, compared to when students are only supplied with blank paper.

For many years, research has suggested that the act of notetaking is beneficial for memory (e.g., Einstein, Morris & Smith, 1985; Fisher & Harris, 1974; Kobayashi, 2005; Locke, 1977; Peper & Mayer, 1978) both at the point of encoding and as an aid for later revision. Di Vesta and Gray (1972) referred these as the encoding and external storage functions, respectively. With regard to the encoding function, there are a variety of cognitive mechanisms that may be causing the memorial benefits. First, notetaking involves *self-generation;* students typically paraphrase, rephrase, or generate extra information in their own words whilst notetaking. Memory experiments dating back to the 1970s have shown that self-generating information (e.g., generating the to-be-remembered target *HOT* from the antonym *cold*) results in better memory performance compared to reading it, a phenomenon referred to as the generation effect (e.g., Slamecka & Graf, 1978). Second, notetaking involves *production*; that is, students must produce information to create notes rather than read silently. Research has shown that information produced either vocally (e.g., MacLeod, Gopie, Hourihan, Neary & Ozubko, 2010), or, more relevant to current discussion, in written form (e.g., Mama & Icht, 2016) is remembered better compared to reading that information silently. Third, notetaking may involve *retrieval practice.* To make coherent notes, students may need to record not just the information currently being taught, but also retrieve information that was stated earlier on. Countless studies have shown retrieving an item from memory following learning improves memory at a final test compared to restudying it (e.g., Roediger & Karpicke, 2006).

Despite these concerns about the deleterious effects of slide handouts, there is some evidence suggesting that they are actually good for learning. Marsh and Sink (2010) conducted two experiments comparing test performance for longhand notetaking vs handout annotation. Participants viewed a pre-recorded lecture in either a "handout-present" (printed slides *and* blank paper) or "handout-absent" condition (blank paper only) and were tested either immediately or following delay. They found that for both immediate and delayed testing, participants performed equally across both conditions. They therefore concluded that the presence of a handout allows for more efficient encoding, given that less effort is required to annotate a slide handout than to make comprehensive longhand notes. Marsh and Sink also added a restudy condition to their research, in which half of all the lecture material was restudied, either for unlimited time (Experiments 1 and 2) or for 2-min (Experiment 2). They found that, following delay, restudying the material improved test performance equally for both the handout-present and handout-absent conditions, but that those in the handout-present condition spent less time reviewing the material than those in the handout-absent condition. These findings imply that the provision of a handout is advantageous, as less time was needed in restudy to achieve the same outcome as those who were not provided with a handout

Similarly, Babb and Ross (2009) used real university lectures, delivered across two semesters to investigate the benefit of handouts on test performance at end-of-semester testing. Both courses were taught in each semester. In semester 1, one course had slide handouts available prior to the lecture and the other only had handouts available after the lecture. This assignment of course to handout availability condition was reversed in the second semester. Babb and Ross found no difference in test scores regardless of encoding method used.

Both Marsh and Sink's (2010) and Babb and Ross' (2009) research finds benefit in providing a slide handout. However, both studies have confounds which make drawing conclusions difficult. The laboratory environment of Marsh and Sink meant that extraneous variables (such as restudy time) were controlled. However, they provided paper for longhand notes in both conditions make it impossible to conclude which of the two strategies tested (handout annotation or longhand notetaking) is more effective. Whilst these findings indicate having both a handout and making longhand notes is advantageous, it does not provide insight into which of these strategies is providing the advantage. In terms of Babb and Ross, there are two main confounds within this research. Firstly, the only measurable variable that was manipulated was whether the handouts were present prior to the lecture. The naturalistic setting of the research meant that students themselves made decisions as to whether to download and print slide handouts in the condition when they were available before the lecture – indeed, 18% stated that even when slides were available, they did not download them until after the lecture and a further 19% said that they never downloaded the slides. With such a high proportion of the class not adhering to the experimental conditions of the research, it is difficult to draw firm conclusions. Secondly, as discussed previously, it is not possible to isolate the effect of encoding on later testing when many factors, particularly restudy between encoding and testing are likely to have a significant influence on the outcome.

Coria and Higham (2018a) controlled for these confound of both of the above studies. They conducted two experiments in which several groups of participants watched two medical lectures and their memory for the lecture material was tested both immediately and following a week-long delay. Participants were assigned to groups that varied according to the activity they were required to perform whilst viewing the lecture. Three groups are of particular relevance to the current discussion. Specifically, the notetaking group wrote longhand notes on blank paper, the annotation group was provided with slide handouts which they annotated, and a control group just watched the lecture with no learning aids. In both experiments, participants in the notetaking group outperformed both other two groups on both the immediate and delayed test. Moreover, performance in the annotation and control groups did not differ. Coria and Higham concluded that (a) the provision of a slide handout puts students at a disadvantage at the point of encoding compared to providing blank paper because it blocks longhand notetaking and (b) annotating slides does not confer any learning benefits at all. Coria and Higham (2018b) found concordant results in a series of follow-up studies.

Whilst Coria and Higham (2018a) is novel in that it is the first research to conclusively demonstrate such a low efficacy of slide handout annotating during encoding, the research is limited in other ways. Firstly, the interval between immediate and delayed testing was one week. A common time frame for a delayed test post-learning range from 24-hrs post learning (for example, Bui, Myerson & Hale, 2013; Chan, McDermott & Roediger, 2006) to 7 days post-learning (for example, Roediger & Karpicke, 2006; Vaughn & Rawson, 2011). However, this may not be an educationally relevant interval. In a university environment, the delay between learning and testing can be considerably longer, as most university courses take place over a semester of two months or more. The second limitation in Coria and Higham's (2018a) research is that there was no restudy condition within the design of their experiments. Whilst the inclusion of at least some no-restudy conditions is necessary to isolate the benefits of the encoding strategy, students studying at university would typically have the opportunity to restudy lecture material prior to an exam. There is potential that any encoding benefits could be diminished or even eliminated by either increasing the retention interval between learning and testing, or adding a restudy opportunity.

We are unaware of any research into notetaking conducted in a laboratory setting that has looked at longer time intervals between learning and testing. However, naturalistic studies using university courses tend to have longer delays prior to testing. Weatherley, Grabe and Arthur (2003) conducted a study using two university cohorts. Both cohorts sat the same course but one was provided access to lecture handouts to print prior to the lecture for the entire semester, and the other was not given access to handouts. The students were tested using various methods and on varying intervals across the semester, but were also tested in a final test at the end of the semester. Results showed that the cohort who had access to the lecture slides showed significantly poorer performance at testing than those who did not have access. This research suggests that the benefit seen for longhand notetaking in Coria and Higham (2018a) should persist across a longer delay. However, there are confounds associated with the naturalistic design, such as a lack of total compliance with encoding condition amongst participants and lack of separation between encoding and restudy, as discussed previously. Furthermore, laboratory research into other learning strategies has indicated that the effects of how material is learned persist over a longer retention interval. For example, Butler and Roediger (2007) replicated the testing effect in a simulated classroom setting and found that the benefits associated with retrieval practice prior to final testing persist following a month-long delay.

As discussed previously, the second limitation of Coria and Higham (2018a) is that they did not provide the participants with a restudy opportunity. This is an important aspect of university life, and one that should be examined before any conclusions are drawn regarding the benefits associated with notetaking during lectures. The interaction between other learning strategies and restudy has shown that restudy attenuates early benefits found during encoding; Rawson and Dunlosky (2013) found that any benefits of spacing during learning are attenuated by restudy, and Vaughn, Dunlosky and Rawson (2016) and Storm, Friedman, Murayama and Bjork (2014) both found that restudy prior to a final test overrides any effects of retrieval practice attempts between learning and testing. This could indicate that engaging in restudy is more beneficial than any other strategy for learning and wipes out any previously seen benefit for that learning strategy.

Indeed, Kierwa et al. (1991) investigated the storage and encoding functions of notetaking across three conditions; notetaking with no review, notetaking and review, and absence from lecture and review. As expected, participants who engaged in notetaking and review showed the best performance at later testing. However, participants who did not attend the lecture, but reviewed the material also out-performed those who engaged in notetaking with no review. This implies that, whilst notetaking is beneficial at encoding, it is less beneficial than restudy. This finding could be interpreted as demonstrating that regardless of the encoding strategy, restudy will provide more benefit since participants were able to outperform notetakers without even attending the lecture! However, since this research was conducted prior to prevalent powerpoint lectures and slide handouts, it does not provide any indication as to whether the benefits seen from notetaking compared to slide annoataion in the absence of restudy are entirely lost, or whether longhand notetakers still out perform their slide annotating counterparts, regardless of restudy. Furthermore, the above research focuses on test performance following restudy, and does not take the potential saving of reduced study time into account

Research dating back to Ebbinghaus (1885) discussed the idea of savings, when relearning information. For example, in successive relearning trials (see Bahrick, 1979), time taken to reach criterion (i.e., the correct answer) during restudy is less than the time taken during initial learning. In this vein, we could expect that, as Coria and Higham's (2018a) experiments showed that more information was learned during encoding for longhand notetakers (evidenced by improved memory performance at testing), students engaging in longhand notetaking would need less time at restudy than those who annotated a slide handout. This is something we plan to investigate.

Given that none of the above research adequately answers the question of whether the encoding benefits of notetaking would a) last over educationally relevant intervals or (b) would still exist following restudy, we plan to investigate these two issues in the current research.

Experimental Overview

The aim of this research was to replicate the findings of Coria and Higham (2018a), but to provide a more educationally relevant context, by increasing the retention interval between encoding and testing, and by allowing controlled restudy.

In Experiment 1, we followed the a similar procedure to that used in Coria and Higham (2018a), and continued to exclusively focus on the encoding process whilst increasing the time between encoding and testing. By changing this single aspect, we can ensure that any variations in the findings are directly attributable to this manipulation**.**

Coria and Higham (2018a) took retrospective confidence ratings and found that participants in the longhand notetaking conditions, who showed significantly higher test scores, did not show higher retrospective confidence ratings than slide annotators or controls. In this experiment, we included retrospective confidence for test answers to investigate whether this metacognitive unawareness persists following a longer delay. Research into retrospective confidence ratings for eyewitness reports following several weeks has showed that participants have reduced metacognitive awareness (i.e., they were less able to accurately identify if their answer was correct) following the longer retention interval (Sauer, Brewer, Zweck & Weber, 2009). We predicted that our participants would also show this lack of metacogntive awareness.

Participants were assigned to one of three groups: the control group (who passively observed the lecture without any material for notetaking), the annotation group (who were provided with a slide handout to annotate) and the notetaking (NT) group (who were provided with blank paper to make longhand notes). All participants viewed a single lecture in the first phase of the experiment, before returning eight weeks later for testing.

In Experiment 2, we followed the same procedure as Experiment 1, but we dropped confidence ratings and added a restudy phase. We took measures of both test performance and time spent engaging in restudy across the three independent groups. This allowed us to identify whether any benefits from encoding conditions remained following restudy, either in terms of improved test scores or savings from reduced time needed to be spent restudying.

Experiment 1

Method

Participants. In order to determine group sizes, we reviewed the literature with the most similar methodology which was Coria and Higham (2018a) and Coria and Higham (2018b), which both used groups of approximately 20 participants per encoding method. In this experiment, 93 undergraduate students from the University of Southampton took part in the first session of this research in exchange for course credit. The participants were assigned to an encoding group based on the session that they signed up to attend. However, 23 of these participants did not return following the eight-week delay to take part in the second session. As no data beyond age, year of study, gender and previous experience was taken from the first session, any participant who did not return for the second session was dropped from the study and not included in the analyses. Two participants were removed from the sample after they declared on the screening questionnaire that they had previous experience with the lecture topic of in-vitro fertilization (IVF). The remaining 68 undergraduate students (9 males, 59 females, *M* age = 21.32, *SD* age = 6.45) who attended both sessions make up the sample for this experiment and were assigned to the four encoding groups as follows: Control = 27, Annotation = 18, NT = 23. All participants provided test data during phase two; thus all 68 participants contributed accuracy data. However, due to experimental error, four participants in the control group did not provide confidence ratings for their answers. Therefore, for data regarding confidence ratings, the sample sizes were as follows: Control = 23, Annotation = 18, NT = 23.

Design. The design consisted of three independent encoding groups (control, annotation and NT) who were tested following an eight-week delay. The dependent measures were test accuracy (measured by both multiple-choice and cued-recall questions) and confidence for test answers (measured on a scale 0-100). The experiment took place over two sessions. In Phase 1 participants viewed the lecture in their assigned encoding group. In Phase 2, eight weeks later, participants returned to the lab for testing.

The encoding groups differed in the activity that they engaged in whilst viewing the lecture. Participants in the control group were not given any materials and were simply instructed to watch the lecture. Participants in the annotation group were given a pen and a handout of the printed lecture slides exactly as they appeared in the lecture presentation, in a layout of 2 x 3 slides per page. They were instructed to listen and annotate the handout on any occasion that they would do so in a real lecture.

Participants in the NT group were provided with a pen and blank paper. They were also instructed to take notes when they would do so in a real lecture.

The test consisted of 30 questions in total. Each question related to a key point covered in the lecture. Twenty of the questions were multiple-choice (each with five options) and ten were cued-recall questions. The questions on the test appeared in the order in which the relevant material occurred in the lecture, with cued-recall and multiple-choice questions interspersed randomly. All participants completed the same tests with the same question order.

Materials and procedure.

Phase 1 – lecture. The first phase of the experiment took place in a small lecture theatre (maximum capacity = 25) and the experiment was conducted in groups of between three and 18 participants at one time. For each session, all participants were in the same encoding group.

On arrival, participants were seated in the lecture theatre, facing the screen. They were given a screening questionnaire in which they were asked to provide their age, gender, year of study and whether they had previously studied, or been affected by the lecture topic of IVF. Once all participants had completed the questionnaire and were ready to proceed, they received the verbal instructions and material corresponding to their encoding group, outlined above.

The lecture was pre-recorded to ensure that participants in all sessions received the same material, delivered at the same pace and with the same voice intonation in order to maintain experimental control. The lecture was played on a large screen at the front of the lecture theatre. It consisted of lecture slides created in Apple Keynote, designed to mimic those seen in typical university classrooms. All slides were written in bullet-point format and some slides contained supplementary images but the main lecture content was represented by the text on the slides. The slides were accompanied by a pre-recorded audio voiceover conveying information in the style of a lecturer. The same lecture was given to all participants regardless of encoding group, and the lecture ran for 25-min.

The lecture used in this experiment was about in-vitro fertilization (IVF) and the content related to the reasons for needing the treatment, the processes involved in treatment and the prevalence and incidences of the treatment since its initial success. The whole lecture consisted of 10 sections. Each section lasted approximately 2.5-min in length and consisted of three main, high importance messages, or *key points*, all of which were all included in the later test. An example of a section within the lecture was "History and Incidence of IVF" and a key point within that section was that Louise Brown was the first baby to be born following a successful course of IVF in August 1978. Information pertaining to each key point was included in both the slides and audio voiceover although the wording was not necessarily identical. This was done to mimic a true lecture experience.

Following the lecture, participants were provided with a partial debriefing form to remind them of the room and time to return eight weeks later. The partial debriefing form also instructed them to contact the researchers or consult with their doctor if they had any concerns prior to Phase 2 of the experiment. There was no mention of the aim of the study, or of a second test in Phase 2 to prevent participants from researching the topic between sessions.

Phase 2 – Testing. Eight weeks after Phase 1, participants returned to complete the second phase of this experiment, which took place in individual booths in a small learning laboratory. Once participants had arrived and consented to proceed, they were instructed to leave all belongings outside of the room to avoid distractions. Participants were then all tested on the content of the lecture. The test had a cover page of instructions on how the test paper should be completed (accompanied by verbal instructions delivered by the experimenter). For the multiple-choice questions, participants were instructed to select the answer they believed was the most likely to be correct. Examples of both correctly completed multiple-choice and cued-recall questions (unrelated to the lecture material) were included on the cover page. Underneath their chosen answer, participants were asked to indicate their confidence, on a scale of 0-100, that their answer was correct.

For the cued-recall questions, participants were given space to provide an answer alongside the instruction to guess if unsure (i.e., not to leave blanks). A confidence rating was assigned to each answer again, in a space next to their written answer. Correctly answered examples of both multiple-choice and cued-recall questions can be seen in Figure 1.

To encourage optimal performance, there was no time limit on the test, but participants were asked to indicate when they had completed the test so that the experimenter could fully debrief them. If participants did not return for Phase 2, they were sent the debriefing information by email.

Ethics Approval. The study was reviewed and approved by the University Ethics Committee and the Research Governance Office. All participants were consenting adults over 18 years old and all gave informed consent for their data to be included in the study and any publications that follow from it.

Results and Discussion

Test accuracy. A 3 (group: control, annotation, NT) x 2 (test type: multiple choice vs cued recall) mixed-model Analysis of Variance (ANOVA) on the test scores (accuracy) revealed a main effect of test type, $F(1,65) = 84.25$, $p < .001$, $\eta^2 = .56$, $d = 2.28$. Unsurprisingly, participants scored significantly higher when tested using a multiple-choice question format (*M* accuracy = .43, *95% CI* [.40, .46]) compared to when tested using a cued-recall question format (*M* accuracy = .28, *95% CI* [.25, .30]). There was also a main effect of group, $F(2,65) = 5.56$, $p = .006$, $\eta_p^2 = .15$. To investigate this effect of group in more detail, we conducted planned one-way ANOVAs to compute pairwise comparisons between the control group and each of the experimental groups. These comparisons revealed no significant differences between the control group and the annotation group, $F < 1$, but a significant difference between the control group and the NT group, $F(1,48) = 9.48$, $p = .003$, $\eta^2 = .17$, $d = .89$. Finally, we found an interaction between group and test type, $F(2,65) = 3.96$, $p = .024$, $\eta_p^2 = .11$. As shown in Table 1, the interaction was due to a particularly high MCQ test score for participants in the control group, although not significantly, $F(1,43) = 2.96$, $p = .092$, $\eta^2 = .06$. Table 1 shows a breakdown of test scores by MCQ and cued recall questions for all groups.

Confidence.

Table 2 shows a breakdown of confidence ratings assigned to MCQ and cued recall questions for all groups. A 3 (group: control, annotation, NT x 2 (test type: multiple choice vs cued recall) mixed-model Analysis of Variance (ANOVA) on confidence ratings for test answers revealed a main effect of test type, $F(1,61) = 55.92$, $p < .001$, $\eta^2 = .48$, $d = 1.91$. As with accuracy, participants provided higher confidence ratings when tested using a multiple-choice question format (*M* confidence = 45.02, *95% CI* [41.70, 48.34]) compared to when tested using a cued-recall question format (*M* confidence = 32.58 *95% CI* [28.57, 36.59]). No other effects or interactions were found to be significant, largest $F < 1$.

Overall Findings. The test accuracy findings from this experiment show that longhand notetaking was the most beneficial strategy, with lecture-slide annotation providing no benefit above passive observation. The data collected on confidence for test answers also demonstrates that participant confidence ratings were not affected by group, which mirrors the findings of Coria and Higham (2018a). This implies that, at least during testing, students are not metacognitively aware of the advantage provided by longhand notetaking.

Experiment 1 showed that the benefits of notetaking are still evident despite much longer delays between the encoding phase and testing phase than seen in previous similar research, such as Coria and Higham (2018a). However, we were also interested in investigating whether any restudying that is likely to happen during this delay period would have an impact on the benefits provided by notetaking in the absence of restudying. In Experiment 2, we replaced the measure of retrospective confidence with time spent restudying, and compared it across the three groups, as well as measuring test accuracy.

Experiment 2

Method

Participants. As with Experiment 1, group sizes were determined by previous research using similar methodology by Coria and Higham (2018a) and Coria and Higham (2018b) whose encoding groups consisted of approximately 20 participants.

For this experiment, 120 undergraduate students from our University took part in the first session of this research in exchange for course credit. The participants were assigned to an encoding group based on the session that they signed up to attend. However, 53 of these participants did not return following the eight-week delay to take part in the second session. This high level of attrition is likely because of the long delay of eight weeks between initial learning and testing and restudy, which could result in participants forgetting about the experiment, or obtaining all necessary course credit prior to the second session. As no data beyond age, year of study, gender and previous experience was taken from the first session, any participant who did not return for the second session was not included in the analyses. Therefore, the remaining 67 undergraduate students (9 males, 56 females, *M* age = 19.37, *SD* age = 2.39) who attended both sessions were assigned to the three encoding groups as follows: Control = 22, Annotation = 20, NT = 25.

Design. The design for this experiment was the same as in Experiment 1, except that there was a revision stage added prior to testing.

To revise the material prior to testing, all participants were given a new handout identical to that given to those in the annotation group during the lecture and were instructed to restudy the material by reading it, without writing or making any notes on the handouts. The handout contained no notes or annotations written by any students. Although we are aware that students would typically use more material than a printed lecture handout during restudy, we chose this strategy in order to ensure that the material given to the participants only differed within the groups at the point of encoding, so that any variation in performance could be attributed to how the material encoded. Furthermore, providing a handout to all groups regardless of encoding condition offered them the opportunity to revise all of the material that featured in the final test. If the notetaking group had been provided only with their notes and key testing points were missing from those notes, then participants would not have had the opportunity to restudy the missing key points. Providing the notetaking group with their longhand notes *and* a handout could have provided the group with an advantage as their notes could have provided further cues.

The test consisted of 25 questions in total. Each question related to a different key point covered in the lecture. Fifteen of the questions were multiple-choice (each with five options) and ten were cued-recall questions. The questions on the test appeared in the order in which the relevant material occurred in the lecture, with cued-recall and multiple-choice questions interspersed randomly. All participants completed the same tests with the same question order.

Materials and procedure.

Phase 1 – lecture. The first phase of the experiment took place in a medium sized lecture theatre (maximum capacity = 75) and the experiment was conducted in groups of between three and 12 participants at one time. For each session, all participants were in the same encoding group.

On arrival, participants were seated in the front three rows of the lecture theatre, with at least one empty seat either side of them. They were given a screening questionnaire in which they were asked to provide their age, gender, year of study and whether they had previously studied, or been affected by the lecture topic, the hospital superbug MRSA. Once all participants had completed the questionnaire and were ready to proceed, they received the verbal instructions and material corresponding to their encoding group, outlined previously.

The pre-recorded lecture, which was about the hospital superbug MRSA was a modified version of the lecture used in Coria and Higham (2018b). The content and the voiceover remained the same but the total length of the lecture was reduced by 5-min. This was because Coria and Higham's study included pauses for participants to rate their level of task-unrelated thoughts, a measure not used in this experiment. Thus, these pauses between each lecture section were removed. The lecture was played on to a large screen at the front of the lecture theatre. It consisted of lecture slides created in Apple Keynote, designed to mimic those seen in typical university classrooms. All slides contained full sentences (or bullet points) of text and some slides contained supplementary images but the main lecture content was represented by the text on the slides. The slides were accompanied by a pre-recorded audio voiceover conveying information in the style of a lecturer. The lecture was pre-recorded to ensure that participants in all sessions received the same material, delivered at the same pace and with the same voice intonation in order to maintain experimental control.

Following the lecture, participants were provided with a partial debriefing form to remind them of the room and time to return eight weeks later. The partial debriefing form also instructed them to contact the researchers or consult with their doctor if they had any concerns prior to Phase 2 of the experiment. There was no mention of the aim of the study or a second test in Phase 2 to prevent participants from researching the topic between sessions.

Phase 2 – restudy and testing. Eight weeks after Phase 1, participants returned to complete the second phase of this experiment, which took place in individual soundproof booths in a small learning laboratory. Once participants had arrived and consented to proceed, they were instructed to leave all belongings outside of the room to avoid distractions during the restudy phase. Participants were then all provided with a lecture handout, which was identical in appearance and content to the one given to the annotation group during the lecture. All participants received a new handout, void of annotations or markings, even if they had already previously annotated a handout during the lecture. Participants were then instructed that there would be a test and that they should revise the handout content for as long as they felt was necessary until they felt that they were familiar with the material, and that they could perform well at testing. Participants were instructed that as soon as they felt that they had restudied to their satisfaction that they could call the instructor and proceed with the next task. During the restudy session, participants were not provided with any writing material, and revised by reading the slides.

Following completion of their restudy period, participants engaged in a 10-min filler task. For this task, participants were asked to write down (on a sheet of provided paper), as many alternative (i.e., unintended) uses for eight common household items, such as a hairbrush. This was kept to exactly 10-min per group to ensure an equal interval between the end of the restudy period and testing.

Following the filler task, participants were tested on the content of the lecture. The test had two cover pages of instructions on how the test paper should be completed (accompanied by verbal instructions delivered by the experimenter). For the multiple-choice questions, participants were instructed to provide a rank for each of the five answer options that appeared in a horizontal grid, with “1” vs “5” representing the answer they believed was the most vs the least likely to be correct, respectively.

Examples of both correctly completed multiple-choice and cued-recall questions (unrelated to the lecture material) were included on the cover page.

For the cued-recall questions, participants were given space to provide an answer alongside the instruction to guess if unsure (i.e., not to leave blanks). Correctly answered examples of both multiple-choice and cued-recall questions can be seen in Figure 2.

To encourage optimal performance, there was no time limit on the test, but participants were asked to indicate when they had completed the test so that the experimenter could fully debrief them. If participants did not return for Phase 2, they were sent the debriefing information by email.

Ethics Approval. The study was reviewed and approved by the University Ethics Committee and the Research Governance Office. All participants were consenting adults over 18 years old and all gave informed consent for their data to be included in the study and any publications that follow from it.

Results and Discussion

Test accuracy. As discussed previously, participants ranked the options provided in the multiple-choice questions based on the perceived plausibility of each option. A score was provided for each question depending on the rank that had been assigned to the correct answer. Correct answers ranked 1, 2, 3, 4 and 5 were assigned 1, 0.8, 0.6, 0.4, 0.2, and 0 points, respectively. Initial analyses were conducted using this scoring system for all answers relating to multiple-choice questions. Whilst this method of scoring was useful to ensure participants had considered all of the multiple-choice alternatives, it led to ceiling effects. Therefore, we rescored all of the answers to multiple-choice questions on a correct-or-incorrect basis, assigning 1 point if the correct answer was assigned to rank 1, and zero if it was assigned any other rank.

Table 3 shows a breakdown for test accuracy for each group by test type. A 3 (group: control, annotation, NT) x 2 (test type: multiple choice vs cued recall) mixed-model Analysis of Variance (ANOVA) on the test scores (accuracy) revealed a main effect of test type, $F(1,64) = 52.40$, $p < .001$, $\eta^2 = .45$, $d = 1.81$. Unsurprisingly, participants scored significantly higher when tested using a multiple-choice question format ($M = .81$, *95% CI* [.77, .85]) compared to when tested using a cued-recall question format ($M = .67$, *95% CI* [.62, .72]). No other main effect or interaction was significant from the analysis of accuracy, largest $F(3,64) = 2.92$, $p = .06$, $\eta_p^2 = .08$. This result refers to the main effect of group. Participants showed a marginal effect across the groups with regards to test accuracy, although this was not statistically significant. This result implies that the beneficial effect of longhand notetaking (seen in Coria & Higham 2018a and Coria & Higham 2018b) during lectures appears to be mitigated by restudy opportunities.

Time spent restudying. The mean amount of time spent revising (in minutes) is shown in Figure 3. A one-way, between-subjects ANOVA with three levels (group: control, annotation, NT) on time spent revising prior to testing revealed a main effect of group, $F(2,64) = 8.07$, $p = .001$, $\eta_p^2 = .20$. To investigate this effect of group in more detail, we conducted one-way ANOVAs to compute planned pairwise comparisons between the control group and each experimental group. These comparisons revealed no significant differences between the control group and the annotation group, $F < 1$, but a significant difference between the control group and the NT group, $F(1,45) = 13.66$, $p = .001$, $\eta^2 = .23$, $d = 1.10$.

Overall Findings. These findings indicate that, whilst participants' scores were not influenced by their encoding group following restudy, the amount of time that participants chose to spend restudying the material was reduced for participants in the notetaking group compared to controls. This could be considered surprising, given that participants in the notetaking group had not previously seen a printed handout, thus could be expected to require longer to familiarise themselves with it. To our knowledge, these findings are the first to demonstrate a saving during restudy for students who engage in longhand notetaking. Due to the increased information learned during encoding, the time needed to be spent restudying is reduced.

General Discussion

The findings from our experiments provide evidence that the benefits conferred by longhand notetaking during encoding persist across both a long 8-week retention interval between study and testing both with (Experiment 2) and without (Experiment 1) a restudy opportunity.

Experiment 1 showed that, although the delay between encoding and testing was considerably longer compared to similar previous research (such as Coria & Higham, 2018a and Coria & Higham, 2018b) in line with timings found in a semester-long university course, longhand notetakers continue to outperform slide-handout annotators and passive observers. However, confidence ratings in Experiment 1 showed a lack of metacognitive awareness of this benefit. This research demonstrates that the benefits associated with longhand notetaking persist over time in the same way that other beneficial cognitive phenomena such as the testing effect (Butler & Roediger, 2007) have endured across long time intervals between learning and final testing.

At first glance, the accuracy findings from Experiment 2 appear contradictory to findings in Experiment 1, as well as Coria and Higham (2018a) and Coria and Higham (2018b). As discussed above, Rawson and Dunlosky (2013), amongst others, found that any benefit in the encoding activity, in this case, longhand notetaking, was eliminated by the introduction of a restudy phase, and test performance did not differ across the three encoding groups. However, the time spent in the restudy phase was significantly lower for the longhand notetakers than the other two groups. This shows another benefit of longhand notetaking as less time was needed in restudy to obtain the same result.

As we have seen in Experiment 1 (also Coria & Higham, 2018a; Coria & Higham 2018b), participants in the notetaking group(s) demonstrate increased learning compared to slide annotators and controls through higher test scores. Because of this, we believe that the reduction in restudy time seen for longhand notetakers in Experiment 2 was because they had learned more information than the other groups during encoding. If information has been learned during encoding, a student will spend less time covering that material during restudy. Thus, the more information learned during encoding, the shorter the restudy session.

Whilst our results support the majority of the findings regarding the efficacy of notetaking (such as Coria & Higham, 2018a; Coria& Higham, 2018b; Kobayashi, 2005), they are contradictory to those of Marsh and Sink (2010), both in terms of accuracy and time spent in restudy. Marsh and Sink found no difference in test scores between participants who were provided with a slide handout and those who were not, and that participants provided with a handout needed less restudy time to prepare for a final test. However, the variations in encoding conditions between our experiment and that of Marsh and Sink make direct comparison difficult. Specifically, their “handout-present” condition, and our annotation condition involve different encoding activities. Marsh and Sink provided participants with blank paper as well as a handout, so participants tended to also make notes on blank paper. This could lead to participants engaging in the same processes as those beneficial in longhand notetaking in the absence of a handout (i.e., production, generation). Thus, it is unknown if the handout provides any benefit, or indeed any increased efficiency.

University lecturers may be interested to consider our findings that show that the activity of longhand notetaking has proven to be beneficial even after increased time-delay periods, and has extended benefits beyond that of test accuracy. It is possible that some lecturers assume that the format in which information is provided in lectures is not important because students will engage in restudy prior to examination. Whilst our findings show that restudy does allow for students who did not encode as much information during the lecture to “catch up”, this will cost more time. Thus, slide annotation is a less efficient method of learning compared to longhand notetaking. Also, students will not necessarily restudy all of the material provided in a lecture, particularly if their time available for restudy is constrained. The findings of Experiment 1 show that, even after a longer delay period of eight weeks, material not restudied was more likely to be remembered by those who engaged in longhand notetaking during the lecture. The strategy of longhand notetaking is beneficial whether the material is restudied or not.

Whilst our research demonstrates that longhand notetaking is more beneficial than annotating a complete lecture slide handout, we cannot claim that this is necessarily true for all students. There are bound to be interpersonal factors affecting the success of notetaking as a strategy. For example, Williams and Eggert (2002) suggested that a person’s listening, cognitive processing and ability to record the lecture in notes all contribute to a persons’ notetaking skills, which in turn predicts performance. Furthermore, we have not yet considered any motivational impact that a potential removal of slide handouts would have. Whilst it would appear to be an external motivating factor to students that longhand notetaking can lead to less time spent in restudy, research by Wongkietkachorn, Prakconsukapan and Wangsaturaka (2014) showed that concentration decreased and class-skipping increased when handouts were not present.

It is necessary to emphasise that our research only suggests that the provision of *complete* slide handouts can be detrimental for learning. It is known that lecturers sometimes produce “skeletal notes” to address complex topics, and some research (e.g., Kam et al., 2005), suggests that providing some lecture notes in advance of the lecture can increase student capability and understanding of information. As technology advances, we predict that the use of annotation of lecture notes on computer tablet and smartphone will increase, and current research has showed mixed findings; some (e.g., Grabe & Christopherson, 2005) has found benefit in engaging with technology during lectures, whereas Mueller and Oppenheimer (2014) have suggested that using computers to type lecture notes merely makes learning more fluent, reducing the desirable difficulty (Bjork & Bjork, 2011) of notetaking.

Finally, it is important to acknowledge that almost all university courses require the student to engage in their own research and independent learning. It would be erroneous to suggest that a strategy that will improve memory for lectures will automatically lead to better grades at university. However, lecturers are responsible for providing the basis on which students will conduct their own research and encouraging strategies that promote learning during lectures will increase a student’s chance of high performance at university.

Paper 3 – Tables

Table 1

Mean *(Standard Deviation)* [±95% Confidence Limits] Test Scores as a Function of Test Type and Group.

Group	Test Type		
	MCQ	CR	Total
Control	.43 *(.10)* [.39, .47]	.22 *(.12)* [.17, .26]	.33 *(.09)* [.29, .36]
Annotation	.38 *(.10)* [.33, .43]	.28 *(.12)* [.22, .33]	.33 *(.09)* [.28, .37]
NT	.47 *(.11)* [.43, .52]	.33 *(.11)* [.28, .38]	.40 *(.09)* [.36, .44]

Note: MCQ = Multiple Choice Questions, CR = Cued-recall questions, NT = Notetaking

Table 2

Mean (Standard Deviation) [±95% Confidence Limits] Confidence (%) Ratings as a Function of Test Type and Group

	Test Type		
Group	MCQ	CR	Total
Control	44.37 *(9.53)* [38.86, 49.87]	33.33 *(11.94)* [26.29, 39.97]	38.85 *(8.60)* [33.41, 44.29]
Annotation	43.54 *(12.73)* [37.32, 49.76]	31.22 *(18.06)* [23.71, 38.73]	37.38 *(13.22)* [31.23, 43.53]
NT	47.14 *(16.35)* [41.64, 52.65]	33.20 *(17.58)* [26.56, 39.84]	40.17 *(15.49)* [34.73, 45.61]

Note: MCQ = Multiple Choice Questions, CR = Cued-recall questions, NT = Notetaking

Table 3

Mean *(Standard Deviation)* [±95% Confidence Limits] Accuracy on the Memory Test in Experiment 2 as a Function of Group and Test Type

Group	Test Type		
	MCQ	CR	Total
Control	.80 *(.20)*	.61 *(.23)*	.71 *(.21)*
	[.73, .86]	[.53, .70]	[.64, .78]
Annotation	.82 (.13)	.67 *(.20)*	.74 *(.15)*
	[.75, .89]	[.57, .76]	[.67, .82]
NT	.80 *(.13)*	.72 *(.19)*	.76 *(.15)*
	[.74.87]	[.64, .80]	[.70, .83]

Note: MCQ = Multiple Choice Questions, CR = Cued-recall questions, NT = Notetaking

Paper 3 – Figures

Which of the following is NOT a predictor of IVF success?

	A	B	C	D	E
Options:	Weight of male	Age of female	Weight of female	Caffeine intake	Semen Quality from donor sperm
Answer	X				
Confidence:	75%				

What year was the first successful IVF baby born?

Answer	Confidence
1978	80%

Figure 1. Sample multiple-choice (top panel) and cued-recall (bottom panel) test questions, both with confidence ratings (0-100%). Both questions were answered correctly.

1) Which method is used to detect MRSA in people?

	A	B	C	D	E
Answers:	Blood Test	MRI	Throat Swab	Gait Test	Nasal Swab
Rank:	3	4	2	5	1

1) Romania has the highest rate of infection from MRSA, which European country has the lowest rate of infection?

Answer
Iceland

Figure 2. Sample multiple-choice (top panel) and cued-recall (bottom panel) test questions. Both questions were answered correctly.

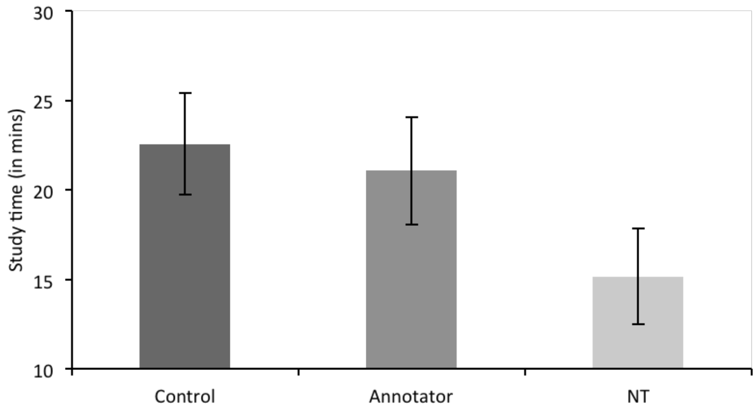


Figure 3. Mean amount of time spent revising (in minutes). NT = notetaking. Error bars indicate 95% confidence intervals.

General Discussion

Throughout this research, the aim was to determine whether longhand notetaking was more beneficial for learning at the point of encoding (during lectures) than annotating a lecture slide handout. Through six experiments over three papers, we found an unambiguous positive effect of notetaking in lectures compared to lecture side handout annotation, and compared to our control group who passively observed the lecture. Our main dependent variable throughout this research was memory at testing, which we measured by test score. However, we also took several other measures described in the three papers to give us indications of any potential further benefits that longhand notetaking could provide for students

Using Jenkins' Tetrahedral Model to Examine any Limitations of Longhand Notetaking

Beyond establishing the existence of a benefit for longhand notetaking during lectures, we used Jenkins' (1979) tetrahedral model of learning to design experiments with manipulations which could investigate the extent of this benefit, and identify aspects of a lecture that could affect whether longhand notetaking is a better strategy for learning than lecture slide handout annotation. Jenkins' model is comprised of four key factors; events (occurances during learning and the environment in which the material is learned), criterion task (how the learning is measured), subjects (the interpersonal characteristics of the participants), and activities (how learning is undertaken).

Within all of our six experiments across three papers, we manipulated the activities factor by controlling how participants encoded the information during the lecture. For each experiment, we included *at least* three encoding groups; a control group, who passively observed the lecture without access to a pen or paper to make any form of notes, an annotation group who were given a lecture slide handout and a pen to annotate the slides, and a notetaking group, who were given a pen and blank paper to make longhand notes. However, in several experiments, we included further groups. These, and our other manipulations in the model are discussed below.

Paper 1. In the first experiment of Paper 1, we manipulated the events factor by including a between-subject fluency variable; we hypothesized that a disfluent lecture (with an irregular pace, inappropriate intonation and corrected errors) could be confusing for a student making longhand notes. However, as found in Carpenter, Wilford, Kornell and Mullaney (2013) and Carpenter, Mickes, Rahman and Fernandez (2016), test scores did not differ across the fluent and disfluent lectures, and participants making longhand notes scored equally across fluent and disfluent lectures.

We also included two different criterion tasks: an immediate (a test completed on the same day as the lecture, following on from a ten-minute filler task), and a delayed test (completed a week later). We included two time-periods in order to calculate the amount of forgetting that occured between immediate and delayed testing. Research dating back to Atkinson and Shiffrin (1968) showed that better learning leads to increased retrieval strength, which is more resistant to forgetting and thus is remembered for longer. This is necessary when researching strategies to apply to university learning, as testing is rarely immediately following the learning session, and the aim of all University courses is to provide transferrable information, which can aid the students in future careers. Therefore, in order for a learning strategy to be deemed effective, it must be resistant to the effects of decay over time, and must persist into the longer term.

As part of Experiment 1, we also included a fourth encoding group in which we asked participants to take notes for an absent friend who was unable to attend the lecture. We hypothesized that participants could engage with the encoding process in a different way, if they were making notes for someone else, and that this could encourage them to take more complete notes, which could lead in turn to more learning. Research by Nestojko, Bui, Kornell and Bjork (2014) found that participants who were expected to teach the material that they were learning showed improved performance at testing, and more organized notes, thus we expected that the requirement to create material suitable for an absent friend would have the same effect on motivation. This was the only manipulation of the subject factor of Jenkins' (1979) model included in our research. However, we found that whilst participants who made notes for a friend out-performed controls and annotators, they performed equally with the regular notetaking group, and this encoding group was dropped from all further experiments.

Following the results of Experiment 1, we changed the criterion task for measuring learning in Experiment 2 from multiple choice to cued recall. This was manipulated to remove any aspect of guesswork within multiple-choice learning, which does not represent genuine learning, and allowed us to distinguish between participants who were selecting multiple choice answers based on recognition, a process which requires fewer cues to lead to retrieval than recall (Gillund & Shiffrin, 1984). Furthermore, Kobayashi (2005) found that the effect of notetaking was more apparent for recall tests than recognition-based tests (such as multiple choice). Our findings showed once again that participants in the notetaking encoding group showed higher test scores both following immediate and delayed cued-recall tests when compared to the annotation and control groups.

To address the events factor in Experiment 2, we changed the material from medical topics (tuberculosis and influenza) in Experiment 1, to natural science topics (ozone layer destruction and acid rain) in order to remove any possibility of the lecture material in Experiment 1 contributing to the results found. Furthermore, in Experiment 2, we added two further within-subjects variables within the events factor of the model: lecture speed and lecture content.

In order to manipulate lecture speed, participants viewed two lectures; one at a regular pace and one presented 30% faster. We included this variable as it seemed plausible that the benefits associated with notetaking could be moderated if the pace became too rapid for effective notetaking. This is because notetaking, being a more cognitively difficult task than slide annotation (Piolat, Olive & Kellogg, 2005), is more time-consuming and thus the increased speed could place more demand on working memory. This could mean that the task of notetaking becomes ineffective and a student is unable to effectively attend to the lecture and make longhand notes. This theory is supported by research by Kobayashi (2005) and Kierwa (1985) who suggested that presentation speed could be a moderator to the benefits of notetaking. However, research showed that there was no difference in test performance across the encoding groups - longhand notetakers out-performed slide annotators and the control group for both the regular-speed and faster-speed lectures.

The manipulation that we added regarding lecture content refers to the type of information included in the lectures. For each lecture, we broke the content down into two content groups: fact and concept. We defined concepts as processes that involved a series of steps to complete an idea unit, and facts as single idea units which could be understood without relying on any further information within the lecture. We chose to investigate this distinction to determine whether the concept-based lecture content could cause problems for notetakers, if they were unable to adequately note down all of the steps in order to be able to understand the concept as a whole. We also theorized that concept based material would be particularly challenging for notetakers in a faster paced lecture, due to a progressive cognitive overload. However, whilst we found that across all groups, performance was reduced for concept-based material in the faster paced lecture, this was not more apparent for notetakers. Furthermore, notetakers scored better on both fact- and concept-based material than annotators and the control group.

The results of Paper 1 showed that none of the variables that we manipulated (described above) moderated the benefits associated with longhand notetaking in terms of performance on testing. Lecture slide annotators' performance did not differ from passive observers.

Paper 2. In our second paper, we focussed on manipulating the activities factor and added three extra encoding groups. As well as the control group, the annotation group and the notetaking group, we added a second notetaking condition in which participants were instructed to write as much of the material as they could, verbatim from the lecture. This group was added to determine whether certain types of notetaking are more beneficial than others, as we theorised that the success of the strategy of notetaking is determined by how generative the activity is (Armbruster, 2000). We predicted that participants making verbatim notes, who engaged in less generation would remember less information at testing than those taking regular notes. Our results showed that participants in the verbatim notetaking group initially showed better performance than the regular notetaking group, outperforming them on immediate test. However, by delayed testing, regular notetakers showed significantly higher test scores. This finding replicates the work of Bui, Myerson and Hale (2013) who also found that participants who engaged in verbatim notetaking showed better performance compared to organised notetakers immediately following learning, but following delay, this pattern was reversed.

Following the vast literature demonstrating the efficacy of retrieval practice as a beneficial learning strategy (e.g., Karpicke & Blunt, 2011; McDaniel, Anderson, Derbish, & Morissette, 2007; Roediger & Butler, 2011), we added two further groups who engaged in retrieval practice during encoding. Participants in these conditions were instructed not to write during the lecture, but were provided with a pause at the end of each lecture section in which to note down the key points of the previous section. One of the two encoding groups did not receive any feedback as to whether their key points were correct but the second group received feedback. This is the first research that we are aware of, to actively implement a controlled form of retrieval practice into encoding during initial learning. Previous research (discussed above) investigated retrieval practice as a form of restudy following initial learning.

Results showed that the participants who engaged in retrieval practice in this way, either with or without feedback showed the best performance across all of the groups in this experiment. We expected the group who received feedback to out perform the group who did not, given that corrective feedback is highly beneficial for memory performance (Butler, Karpicke & Roediger, 2007). However, both groups performed equally at both immediate and delayed testing.

In terms of criterion task, we used a mix of multiple-choice and cued-recall in order to identify whether any of the encoding groups led to improved performance in a particular test type. However, our findings persisted across both test types and Experiment 1 identified a new encoding strategy that shows further benefits to longhand notetaking – retrieval practice with or without feedback. However, a potential explanation for these groups showing improved performance is that the amount of exposure to the lecture material was increased. Participants had longer to learn the lecture material as, in order to incorporate pauses into the lecture, the length of the lecture was extended.

To further investigate this, we conducted a second experiment in which we compared another retrieval practice group who received feedback, who spent the longest time viewing the lecture in Experiment 1, with a modified control group, who were also provided with lecture pauses. However, instead of writing the key points, participants in the modified control group were encouraged to think about the material that they had studied in the previous section. The results showed that, at immediate testing, there was no difference between the groups. However, following delay, participants in the retrieval practice with feedback group out performed their control counterparts.

Paper 3

Both Paper 1 and Paper 2 showed that the benefits of longhand notetaking (and some modifications of longhand notetaking) persist across various activities, events and types of criterion. However, the retention interval used between learning and delayed testing was considerably shorter than would be expected during a university course. Thus, we increased the retention interval from one week to eight weeks in order to investigate whether the increased delay affected the efficacy of longhand notetaking. Given that research on retrieval practice showed that the effect persisted over a longer retention interval (Butler & Roediger, 2007), we predicted that longhand notetakers would still out perform slide annotators and controls. Indeed, as with Paper 1 and Paper 2, longhand notetakers showed better memory for testing in a combination of multiple-choice and cued-recall questions.

In Experiment 2, we retained the eight-week delay and incorporated the highly relevant activity of restudy prior to testing. We anticipated that restudy would reverse any benefit of longhand notetaking following encoding. However, we also predicted that, since longhand notetakers in Experiment 1 had demonstrated more learning by scoring higher at delayed testing, they would need less time during restudy. As predicted, we found no difference between the groups following restudy. However, the amount of time spent engaging in restudy was lower for the longhand notetakers than the slide annotators and controls, demonstrating a saving from their encoding condition, and another highly relevant benefit of longhand notetaking.

Overall findings

Each experiment manipulated different variables in Jenkins' (1979) model. Through doing so, we were able to draw three important conclusions as to the efficacy of notetaking.

Firstly, participants engaging in the activity of longhand notetaking during encoding showed higher scores than slide annotators regardless of lecture fluency, lecture speed, lecture content, criterion task and criterion length. Furthermore, slide annotators' performance does not differ from that of passive observers, showing little benefit from this strategy.

Secondly, "longhand notetaking" is a broad definition and the type of notetaking can predict the benefit of the activity. Whilst copying information verbatim proved to be beneficial only in the short-term, engaging in retrieval practice during lectures, either with out without corrective feedback provides further benefit than just writing notes.

Thirdly, whilst the benefits associated with notetaking during encoding are removed following encoding, the amount of time spent in restudy is reduced as participants have learned more material at encoding.

Limitations and Further Research

All of these conclusions support our hypothesis that the act of longhand notetaking during lectures is beneficial for learning. However, there are other factors to consider, that we have not been able to investigate, particularly regarding the subjects factor of Jenkins' (1979) model. It is essential to acknowledge that various individual differences will affect the extent to which longhand notetaking is beneficial. For example, research by Kierwa and Benton (1988) found that information-processing ability is directly related to notetaking, and that the amount of notetaking is directly related to academic performance. Since this research was conducted prior to the presence of lecture slide handouts, it is unknown whether students with lower information-processing ability would benefit more from a slide handout. However, the findings demonstrate that students' notetaking ability can attenuate the benefits provided by the strategy. Furthermore, Peverley et al. (2013) found transcription fluency (the ease with which a student is able to comprehensively write down information) to be the biggest predictor of quality of notes, and that quality of notes was the only predictor of test performance. This is related to student ability. If a student is struggling to attend to the information whilst notetaking, the difficulties associated with notetaking which cannot be overcome and the benefit of notetaking is no longer realised. A potential avenue for further research would be to attempt to identify an optimal point of performance at which notetaking becomes beneficial. This could be achieved by using a within-subjects design of multiple lectures of varying difficulty, and comparing their test scores when notetaking to a control group. I would be predicted that when the lecture became too difficult and notetaking alongside concentrating on the lecture became non-beneficial due to lack of understanding, the pattern of results would reverse and the control group simply attending to the lecture would out-perform the notetaking group.

Other students who could face difficulties with notetaking include those with a condition that affects learning, such as dyslexia (Boyle, 2010; Mortimore & Crozier, 1996), and those studying in a second language (Chamot & Kupper, 1989).

Given the above difficulties with notetaking for certain student groups, a potential alternative solution is for lecturers to provide a printout containing partial information or skeletal notes.

There is no single definition for skeletal notes, but Hartley (1976) describes them as notes where the main ideas are written with space for students to expand the ideas in their own notes. Indeed, Kierwa (2002) suggested that these skeletal notes might offer a "middle ground" (p.72) between notetaking and providing a complete lecture slide printout. However, the definition provided for skeletal notes indicates that the key points from the lecture would be pre-written. Our research has demonstrated that the generation of key points (Paper 2) was a highly effective encoding strategy. Since the key points are the most likely points to be examined on, skeletal notes could be reducing participants' abilities to generate this key information. Our research indicates that the generation of this key information leads to more learning during encoding, thus better test performance both immediately following learning and after delay.

However, there are other forms of partial notes that have been researched and the findings have been inconclusive. Markovits and Weinstein (2018) described their form of guided notes as instructor-prepared information sheets to cue students to respond to key information. This form of guided notes is based around the premise of the generation effect (Slamecka & Graf, 1978) and encourages students' notetaking to be generative. Markovits and Weinstein proposed that guided notes make student notes more organised, which would be beneficial for students who have lower information processing ability or reduced transcription fluency.

Konrad, Joseph and Everleigh (2009) conducted a meta-analysis into the benefits of guided notes and compared eight studies, of varying ages and abilities of students. Although they found an overall positive benefit of using guided notes compared to not using them, the benefits for students in higher education were inconclusive and less apparent than those in school-aged students. Furthermore, it is not clear what the students who were not provided with guided notes were doing (i.e. whether they were passively observing, annotating a complete handout or making longhand notes or another strategy). Thus the extent of the benefit of these guided notes is unclear. Equally, Stark-Wroblewski, Kreiner, Clause, Edelbaum and Ziser (2006) found no significant difference in performance whether participants were given regular lecture slide handouts (with complete content, as were used in our experiment), or guided notes. From this, we could predict that participants with guided notes would perform at approximately the level of our lecture slide annotation group.

Despite inconclusive findings, the principles underlying Markovits and Weinstein's (2018) guided notes concept appear to include lots of the cognitive strategies seen to benefit notetaking in our research, whilst providing a structure and organisation for students who have reduced notetaking skills. Therefore, this idea requires further research. A potential study to compare the outcome of guided notes with our longhand notetaking group, slide annotation group and control group, with restudy controlled to isolate the encoding function, would provide further evidence as to the efficacy of this strategy in line with those discussed in our research.

It is also necessary to acknowledge that, whilst participants in our experiments were not specifically told of there being a test following the lectures that they viewed, they may have anticipated that testing would be used to measure their learning, thus behaved accordingly when viewing the lecture. This would have been particularly apparent for the first two experiments (Paper 1) in which two lectures were immediately followed by tests – students likely predicted the presence of a second test following the second lecture. However, given that all conditions were provided with the same instructions, this does not appear to be a particular limitation to our findings.

Furthermore, in a real lecture environment, students are aware if they are to be tested on the material, so would likely behave in a similar manner. Furthermore, all items included on the test were directly taken from the lecture slides. This was necessary to identify learning from the lectures however, a potential further avenue for research would be to test for transfer of information to items not directly included in the lecture. It is possible that longhand notetaking would lead to improved transfer of information compared to slide annotation, in a similar vein to Butler's (2010) finding that repeated testing promoted transfer compared to restudy.

A final limitation to our research is that our sampling method of opportunity sampling through undergraduate recruitment in exchange for course credit could have influenced our results in terms of motivation, as it would be expected that a more motivated student would sign up to participate. Indeed, Muryama, Pekrun, Lichtenfeld and Vom Hofe (2013) found that motivation, intelligence and the cognitive strategies used during learning (in this case, notetaking) jointly predict achievement over time. Thus this is something that must be considered. It is clear that a more motivated student is more likely to be willing to engage with the strategies which we have found to enhance learning, whilst a non-motivated student could choose not to make notes during a lecture which, in accordance with our findings, would leave them with no further benefit than the control group. As previous research has indicated that students prefer to receive a lecture-slide handout, it could imply that this would motivate students more during a lecture. However, given that our research did not find any advantages associated with annotating a lecture slide handout compared to the passively observing control groups, it does not seem that it is necessary to recommend the use of a lecture slide handout in cases of lesser-motivated students, providing that access to material is provided after the lecture.

Policy Recommendations

Based on our findings, our recommendations to instructors in universities is that encouraging students to take longhand notes during lectures will lead to more learning during encoding, compared to the situation when students are allowed access to complete lecture slide handouts. Whilst it appears beneficial to provide notes to encourage easy learning, providing obstacles for students to overcome forms a desirable difficulty which can lead to deeper learning during the lecture which, in the absence of restudy, leads to improved test performance. Our final paper also shows that the benefits associated with longhand notetaking can time savings during later restudy, with students who engage in longhand notetaking needing less time to restudy material prior to testing – a clear benefit for students.

Furthermore, our second paper demonstrated that providing lecture pauses in which to encourage students to engage in retrieval practice further bolstered memory at later testing. This strategy could be adapted into a lecture to encourage testing during learning, and promote good habits for later restudy. However, it is necessary to acknowledge that lecturers have a limited time period in which to cover all of the necessary information, and providing pauses could lead some of the material being sacrificed due to time constraints, which may not be beneficial for students.

Finally, an important factor to consider, when recommending longhand notetaking, is to provide instructions as to the type of encoding strategy that is beneficial, to encourage good practice amongst students.

For example, our research has shown that verbatim notetaking is less effective than regular notetaking, but this might not be clear to a student who has no experience of university learning, who might believe that writing everything down will help them to learn. Kobayashi (2006) found that how instructions were given on notetaking was one of the moderators of the modest benefit found for this strategy. Therefore, it is inevitable that explaining the benefit of generative notetaking and retrieval practice, alongside alleviating concerns regarding desirable difficulty, will lead to greater understanding by the student of the encoding practices that will lead to better learning during lectures. Another added benefit of this education is that students will likely incorporate these strategies into their restudy habits, which will further improve test performance.

It is necessary to acknowledge that whilst a lecturer can recommend strategies and provide a framework for good encoding practice, it is ultimately the responsibility of the student to adopt such strategies. All higher education courses require a student to engage in independent research and learning, so it is erroneous to imply that the strategies outlined in this paper will automatically equate to better grades at university for each student. However, it is the responsibility of a lecturer to provide a framework on which students can develop their own ideas. University performance is frequently measured by student grades, but the purpose of higher education is to equip students for future careers and life experiences. Encouraging strategies that promote learning will increase a students' potential during their years at university, but will also develop lifelong learning habits to assist in their future endeavours.

Conclusions

The three papers in this research contribute to a growing body of highly topical research aimed at improving student learning and memory through cognitive psychology principles. At a time when universities are rapidly expanding in student numbers, there is increasing pressure to provide techniques that can be used to improve learning to larger-sized student audiences without any financial burden. Our research found a universal benefit to longhand notetaking during lectures, which is a strategy that can be taught to a class of unlimited size and at no financial cost. We therefore recommend that instructors in higher education consider reverting to the strategies used by previous generations of university students, and refrain from providing lecture slide handouts until the students have had the opportunity to encode the material using a pen and blank paper.

List of References

Aiken, E. G., Thomas, G. S., & Shennum, W. A. (1975). Memory for a lecture: Effects of notes, lecture rate, and informational density. *Journal of Educational Psychology*, *67*, 439-444. doi: 10.1037/h0076613

Amare, N. (2006). To slideware or not to slideware: Students experiences with PowerPoint™ vs. Lecture*. Journal of Technical Writing and Communication, 36,* 297–308. doi: 10.2190/03GX-F1HW-VW5M-7DAR

Armbruster, B. B. (2000). Taking notes in lectures. In R. F. Flippo, D.C. Caverly (Eds.) *Handbook of college reading and study strategy research* (pp. 175-199). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes1. In *Psychology of learning and motivation* (Vol. 2, pp. 89-195). Academic Press.

Avval, F. Z., Jarahi, L., Ghazvini, K., & Youssefi, M. (2013). Distribution of Handouts in Undergraduate Class to Create More Effective Educational Environment. *Int J Ed Res*, *12*, 1-6. doi: 10.1556/030.62.2015.2.4

Babb, K. A., & Ross, C. (2009). The timing of online lecture slide availability and its effect on attendance, participation, and exam performance. *Computers & Education*, *52*, 868-881. doi: 10.1016/j.compedu.2008.12.009

Barbetta, P. M., & Skaruppa, C. L. (1995). Looking for a way to improve your behavior analysis lectures? Try guided notes. *The Behavior Analyst*, *18*, 155-160. doi: 10.1007/bf03392701

Barabasz, A. F. (1968). A study of recall and retention of accelerated lecture presentation. *Journal of Communication*, *18*, 283-287. doi: 10.1111/j.1460-

2466.1968.tb00077.x

Bartsch, R. A., & Cobern, K. M. (2003). Effectiveness of PowerPoint presentations in lectures. *Computers & education*, *41*, 77-86. doi: 10.1016/s0360-1315(03)00027-7

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*(, 55. doi:10.1037/0096-3445.127.1.55

Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. *Memory & Cognition, 41*, 897-903. doi: 10.3758/s13421-013-0307-8

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*, 201-210. doi: 10.3758/bf03193441

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56-64). New York: Worth Publishers.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.) *From Learning Processes to Cognitive Processes: Essays in Honor of William K. Estes*. (pp. 35-67). Hillsdale, NJ: Erlbaum.

Bjork, R.A. (1994). Metamemory and memory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (eds.) *Metacognition: Knowing about knowing* (pp.85-205). Cambridge, MA: MIT press.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417-444. doi: 10.1146/annurev-psych-113011-143823

Blanchard, J., & Mikkelson, V. (1987). Underlining performance outcomes in expository text. *Journal of Educational Research*, *80*, 197–201. doi: 10.1080/00220671.1987.10885751

Boyle, J. R. (2010). Note-taking skills of middle school students with and without learning disabilities. *Journal of Learning Disabilities*, *43*, 530-540.doi: 10.1177/0022219410371679

Bretzing, B. H., & Kulhavy, R. W. (1979). Notetaking and depth of processing. *Contemporary Educational Psychology, 4*, 145-153. doi: 10.1016/0361-476x(79)90069-9

Bromage, B. K., & Mayer, R. E. (1986). Quantitative and qualitative effects of repetition on learning from technical text. *Journal of Educational Psychology*, *78*, 271. doi: 10.1037/0022-0663.78.4.271

Bui, D. C., & McDaniel, M. A. (2015). Enhancing learning during lecture note-taking using outlines and illustrative diagrams. *Journal of Applied Research in Memory and Cognition*, *4*, 129-135. doi: 10.1016/j.jarmac.2015.03.002

Bui, D. C., Myerson, J., & Hale, S. (2013). Note-taking with computers: Exploring alternative strategies for improved recall. *Journal of Educational Psychology, 105,* 299-309. doi: 10.1080/09658211.2014.986135

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1118. doi: 10.1037/a0019902

Butler, A. C., & Roediger III, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*, 514-527. doi: 10.1080/09541440701326097

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied, 13*, 273-281. doi: 10.1037/1076-898x.13.4.273

Butler, A. C., & Roediger III, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*, 514-527. doi: 10.1080/09541440701326097

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review*, *7*, 26–48. doi: 10.3758/BF03210724

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1491-1494. doi: 10.1037/0278-7393.27.6.1491

Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science, 21*, 279-283. doi: 10.1177/0963721412452728

Carpenter, S K., Mickes, L., Rahman, S. & Fernandez, C. (2016). The effect of instructor fluency on students' perceptions of instructors, confidence in learning, and actual learning. *Journal of Experimental Psychology: Applied, 22*, 161–172. doi: 10.1037/xap0000077

Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test?. *Psychonomic Bulletin & Review*, *13*, 826-830. doi:10.3758/bf03194004

Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: instructor fluency increases perceptions of learning without increasing actual learning. *Psychonomic Bulletin & Review*, *20*, 1350-1356. doi:10.3758/s13423-013-0442-z

Chan, J. C., McDermott, K. B., & Roediger III, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, *135*, 553. doi: 10.1037/0096-3445.135.4.553

Chamot, A. U., & Kupper, L. (1989). Learning strategies in foreign language instruction. *Foreign language annals*, *22*, 13-22.doi: 10.1111/j.1944-9720.1989.tb03138.x

Chen, P. H. (2013). The effects of college students' in-class and after-class lecture note-taking on academic performance. *The Asia-Pacific Education Researcher*, *22,* 173-180. doi: 10.1080/10494820.2012.705851

Clark, J. (2008). PowerPoint and pedagogy: Maintaining student interest in university lectures. *College teaching*, *56*, 39-44. doi: 10.3200/ctch.56.1.39-46

Cohn, E., Cohn, S., & Bradley, J. (1995). Notetaking, working memory, and learning in principles of economics. *The Journal of Economic Education, 26,* 291–307. doi: 10.2307/1182993

Cook, L. K., & Mayer, R. E. (1983). Reading strategies training for meaningful learning from prose. In Pressley, M., Levin, J. R. (eds.), Cognitive Strategy Research: Educational Applications, Springer-Verlag, New York, pp. 87-126.

Copley, J. (2007). Audio and video podcasts of lectures for campus-based students: production and evaluation of student use. *Innovations in education and teaching international*, *44*, 387-399. doi: 10.1080/14703290701602805

Coria, K.A., & Higham, P.A. (2018). *PowerPointLESS? Annotating Lecture Slide Handouts is a Less Effective Encoding Strategy than Longhand Notetaking*. Manuscript submitted for publication.

Coria, K.A., & Higham, P.A. (2018). *Activities that enhance learning during lectures*. Manuscript submitted for publication.

Coria, K.A., & Higham, P.A. (2018). *Long-term benefits of longhand notetaking on test performance and revision time*. Manuscript in preparation.

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General, 125*, 159–180. doi: 10.1037/0096-3445.125.2.159

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, *11*, 671-684. doi: 10.1016/s0022-5371(72)80001-x

Crawford, F. (1991), Total Quality Management, Committee of Vice-Chancellors and Principals. *Quality Assurance in Education, 3,* 10-21. doi: 10.1108/09684889310044655

Daniels, L. (1999). Introducing technology in the classroom: PowerPoint as a first step. *Journal of Computing in Higher Education, 10*, 42-56. doi: 10.1007/bf02948722

Di Vesta, F. J., & Gray, G. S. (1972). Listening and note taking. *Journal of Educational Psychology, 63*, 8-14. doi: 10.1037/h0032243

Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review, 8,* 155–161.doi: 10.3758/bf03196152

Douglas, J., Douglas, A., & Barnes, B. (2006). Measuring student satisfaction at a UK university. *Quality Assurance in Education, 14*, 251-267. doi: 10.1108/09684880610678568

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*, 4-58. doi: 10.1177/1529100612453266

Ebbinghaus, H. (1885). *Memory. A Contribution to Experimental Psychology*. New York: Teachers College, Columbia University

Einstein, G. O., Morris, J., & Smith, S. (1985). Note-taking, individual differences, and memory for lecture information. *Journal of Educational Psychology, 77*, 522–532. doi: 10.1037/0022-0663.77.5.522

Evans, L. (1998). Preliminary study: Lectures versus PowerPoint. 4.0. Retrieved July 1, 2015.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160. doi: 10.3758/brm.41.4.1149

Fisher, J. L., & Harris, M. B. (1974). Note taking and recall. *The Journal of Educational Research*, *67*, 291-292.10.1080/00220671.1974.10884632

Grabe, M., & Christopherson, K. (2005). Evaluating the advantages and disadvantages of providing lecture notes: The role of internet technology as a delivery system and research tool. *The internet and higher education*, *8*, 291-298. doi: 10.1016/j.iheduc.2005.09.002

Griffiths, L., & Higham, P.A. (2017). Beyond hypercorrection: remembering corrective feedback for low-confidence errors. *Memory*. Advance Online Publication. doi: 10.1080/09658211.2017.1344249

Fazio, L. K., Agarwal, P. K., Marsh, E. J., & Roediger, H. L. I. I. I. (2010). Memorial consequences of multiple-choice testing on immediate and delayed tests. *Memory & Cognition, 38*, 407–418.doi: 10.3758/mc.38.4.407

Fiedler, K., Lachnit, H., Fay, D., & Krug, C. (1992). Mobilization of cognitive resources and the generation effect. *The Quarterly Journal of Experimental Psychology Section A*, *45*(1), 149-171.doi: 10.1080/14640749208401320

Frey, B. A., & Birnbaum, D. J. (2002). *Learners' Perceptions on the Value of PowerPoint in Lectures.* Retrieved from https://files.eric.ed.gov/fulltext/ED467192.pdf

Gagné, R. M. (1972). Domains of learning. *Interchange*, *3*, 1-8. doi: 10.1007/bf02145939

Gorissen, P., Van Bruggen, J., & Jochems, W. (2012, September). *Students and recorded lectures: survey on current use and demands for higher education.* Retrieved from http://dspace.ou.nl/handle/1820/4418

Gier, V. S., & Kreiner, D. S. (2009). Incorporating active learning with PowerPoint-based lectures using content-based questions. *Teaching of Psychology*, *36*, 134-139. doi: 10.1080/00986280902739792

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological review*, *91*, 1. doi: 10.1037//0033-295x.91.1.1

Healy, A. F., Jones, M., Lalchandani, L. A., & Tack, L. A. (2017). Timing of quizzes during learning: Effects on motivation and retention. *Journal of Experimental Psychology: Applied, 23*, 128. doi: 10.1037/xap0000123

Harknett, R. J., & Cobane, C. T. (1997). Introducing instructional technology to international relations. *PS: Political Science & Politics*, *30*, 496-500. doi: 10.1017/S1049096500046722

Hartley, J. (1976). Lecture handouts and student note-taking. *Programmed Learning and Educational Technology*, *13*, 58-64.doi: 10.1080/1355800760130208

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(, 22–34. doi: 10.1037/0278-7393.29.1.22

Hertzog, C., Hines, J. C., & Touron, D. R. (2013). Judgments of learning are influenced by multiple cues in addition to memory for past test accuracy. *Archives of Scientific Psychology*, *1*, 23–32. doi: 10.1037/arc0000003

Holzl, J. (1997). Twelve tips for effective PowerPoint presentations for the technologically challenged. *Medical Teacher*, *19*, 175-179. doi: 10.3109/01421599709019377

Hunt, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review*, *2*, 105–112.doi: 10.3758/bf03214414

Jing, H. G., Szpunar, K. K., & Schacter, D. L. (2016). Interpolated testing influences focused attention and improves integration of information during a video-recorded lecture. *Journal of Experimental Psychology: Applied, 22,* 305 - 318. doi: 10.1037/xap0000087

Kam, M., Wang, J., Iles, A., Tse, E., Chiu, J., Glaser, D., & Canny, J. (2005). Livenotes: a system for cooperative and augmented note-taking in lectures. In *Proceedings of the SIGCHI conference on Human factors in computing systems*(pp. 531-540). ACM.

Jenkins, J. J. (1979). Four points to remember: A tetrahedral model of memory experiments. *Levels of processing in human memory*, 429-446.doi: 10.4324/9781315796192

Kang, S. H., McDermott, K. B., & Roediger III, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19*, 528-558. doi: 10.1080/09541440601056620

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*, 772-775. doi: 10.1016/j.jml.2006.09.004

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, *32*, 1-24.doi: 10.1006/jmla.1993.1001

Kiewra, K. A. (1985). Investigating notetaking and review: A depth of processing alternative. *Educational Psychologist, 20,* 23-32. doi: 10.1126/science.1199327

Kiewra, K. A. (1988). Cognitive aspects of autonomous notetaking: Control processes, learning strategies and prior knowledge. Educational Psychology, 23, 39-56.doi: 10.1207/s15326985ep2301_3

Kiewra, K. A. (2002). How classroom teachers can help students learn and teach them how to learn. *Theory into practice*, *41*, 71-80.doi: 10.1207/s15430421tip4102_3

Kiewra, K. A., DuBois, N. F., Christian, D., McShane, A., Meyerhoffer, M., & Roskelley, D. (1991). Note-taking functions and techniques. *Journal of educational psychology*, *83*, 240. doi: 10.1037//0022-0663.83.2.240

Kobayshi, K. (2005). What limits the encoding effect of note-taking? A meta-analytic examination. *Contemporary Educational Psychology, 30*, 242-262. doi: 10.1016/j.cedpsych.2004.10.001

Kobayashi, K. (2006). Combined Effects of Note-Taking/-Reviewing on Learning and the Enhancement through Interventions: A meta-analytic review. *Educational Psychology*, *26*, 459-477.doi: 10.1080/01443410500342070

Konrad, M., Joseph, L. M., & Eveleigh, E. (2009). A meta-analytic review of guided notes. *Education and Treatment of Children*, 421-444.doi: 10.1353/etc.0.0066

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science*, *19*, 585–592.

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language, 65*, 85-97. doi: 10.1016/j.jml.2011.04.002

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 989-998. doi: 10.1037/a0015729

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science, 22,* 787-794. doi: 10.1177/0956797611407929

Law, P., Perryman, L. A., & Law, A. (2013, December 09)*. Open educational resources for all? Comparing user motivations and characteristics across The Open University's iTunes U channel and OpenLearn platform.* Retrieved from http://oro.open.ac.uk/39102/

Locke, E. A. (1977). An empirical study of lecture note taking among college students. *The journal of educational research*, *71*, 93-99. doi: 10.1080/00220671.1977.10885044

Lowry, R. B. (1999). Electronic presentation of lectures - effect upon student performance. *University Chemistry Education, 3,* 18-21. doi: 10.32469/10355/15784

Lyle, K. B., & Crawford, N. A. (2011). Retrieving essential material at the end of lectures improves performance on statistics exams. *Teaching of Psychology*, *38*, 94-97. doi: 10.1177/0098628311401587

MacLeod, C.M., & Bodner, G.E. (2017). The production effect in memory. *Current Directions in Psychological Science, 26,* 390-395. doi:10.1177/0963721417691356

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, Cognition, 36,* 671-685. doi: 10.1037/a0018785

Mama, Y., & Icht, M. (2016). Auditioning the distinctiveness account: Expanding the production effect to the auditory modality reveals the superiority of writing over vocalising. *Memory, 24*, 98-113. doi: 10.1080/09658211.2014.986135

Markovits, R. A., & Weinstein, Y. (2018). Can cognitive processes help explain the success of instructional techniques recommended by behavior analysts?. *Npj Science of Learning*, *3*, 2. doi: 10.1038/s41539-017-0018-1

Marsh, E. J., & Sink, H. E. (2010). Access to handouts of presentation slides during lecture: Consequences for learning. *Applied Cognitive Psychology*, *24*, 691-706. 10.1002/acp.1579

Matvey, G., Dunlosky, J., & Schwartz, B. (2006). The effects of categorical relatedness on judgements of learning (JOLs). *Memory*, *14*(, 253-261.doi: 10.1080/09658210500216844

Mannes, S. M., & Kintsch, W. (1987). Knowledge organisation and text organisation. *Cognition and instruction, 4,* 91-115.doi: 10.1207/s1532690xci0402_2

Mayer, R. E. (1983). Can you repeat that? Qualitative effects of repetition and advance organizers on learning from science prose. *Journal of Educational Psychology*, *75*, 40. doi: 10.1037//0022-0663.75.1.40

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*, 494-513. doi: 10.1080/09541440701326154

McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology, 16*, 192-201. doi: 10.1016/0361-476x(91)90037-l

Miele, D. B., Finn, B., & Molden, D. C. (2011). Does easily learned mean easily remembered? It depends on your beliefs about intelligence. *Psychological Science, 22,* 320-324. doi: 10.1177/0956797610397954

Mortimore, T., & Crozier, W. R. (2006). Dyslexia and difficulties with study skills in higher education. *Studies in Higher Education*, *31*, 235-251.doi: 10.1080/03075070600572173

Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard advantages of longhand over laptop note taking. *Psychological Science*, *25*, 1159-1168. doi: 10.1177/0956797614524581

Mulligan, N. W. (1998). The role of attention during encoding in implicit and explicit memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *24*, 27. 10.1037//0278-7393.24.1.27

Murphy, T. M., & Cross, V. (2002). Should students get the instructor's lecture notes? *Journal of Biological Education*, *36*, 72-75. doi: 10.1080/00219266.2002.9655804

Murayama, K., Pekrun, R., Lichtenfeld, S., & Vom Hofe, R. (2013). Predicting long-term growth in students' mathematics achievement: The unique contributions of motivation and cognitive strategies. *Child development*, *84*, 1475-1490. doi: 10.1111/cdev.12036

Nelson, T. O., & Dunlosky, J. (1991). When delaying your judgments of learning can improve the accuracy of your metacognitive monitoring: The delayed-JOL effect. *Psychological Science, 2,* 267–270. doi: 10.1111/j.1467-9280.1991.tb00147.x

Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A revised methodology for research on metamemory: Pre-judgment recall and monitoring (PRAM). *Psychological Methods*, *9*, 53–69. doi: 10.1037/1082-989X.9.1.53

Nestojko, J. F., Bui, D. C., Kornell, N., & Bjork, E. L. (2014). Expecting to teach enhances learning and organization of knowledge in free recall of text passages. *Memory & Cognition*, *42*, 1038-1048. doi: 10.3758/s13421-014-0416-z

Ozubko, J. D., Hourihan, K. L., & MacLeod, C. M. (2012). Production benefits learning: The production effect endures and improves memory for text. *Memory*, *20*, 717-727.doi: 10.1080/09658211.2012.699070

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 3–8. doi: 10.1037/ 0278-7393.31.1.3 craw

Pastötter, B., & Bäuml, K. H. T. (2014). Retrieval practice enhances new learning: the forward effect of testing. *Frontiers in Psychology*, *5*, 286.doi: 10.3389/fpsyg.2014.00286

Peper, R. J., and Mayer, R. E. (1978). Note-taking as a generative activity. *Educational Psychology, 70*, 514-522. doi: 10.1037/0022-0663.70.4.514

Perry, T., & Perry, L. A. (1998). University students' attitudes towards multimedia presentations. *British Journal of Educational Technology*, *29*, 375-377. doi: 10.1111/1467-8535.00083

Peverly, S. T., Brobst, K. E., Graham, M., & Shaw, R. (2003). College adults are not good at self-regulation: A study on the relationship of self-regulation, note taking, and test taking*. Journal of Educational Psychology, 95*, 335-346. doi: 10.1037/0022-0663.95.2.335

Peverly, S. T., & Sumowski, J. F. (2012). What variables predict quality of text notes and are text notes related to performance on different types of tests? *Applied Cognitive Psychology, 26,* 104-117. doi: 10.1002/acp.1802

Peverly, S. T., Vekaria, P. C., Reddington, L. A., Sumowski, J. F., Johnson, K. R., & Ramsay, C. M. (2013). The relationship of handwriting speed, working memory, language comprehension and outlines to lecture note-taking and test-taking among college students. *Applied Cognitive Psychology*, *27*, 115-126.doi: 10.1002/acp.2881

Piolat, A., Olive, T., & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology*, *19*, 291-312. doi: 10.1002/acp.1086

Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General*, *143*, 644-667. doi: 10.1037/a0033194

Rawson, K. A., & Dunlosky, J. (2007). Improving students' self-evaluation of learning for key concepts in textbook materials. *European Journal of Cognitive Psychology*, *19*, 559-579. doi: 10.1080/09541440701326022

Rawson, K. A., & Dunlosky, J. (2012). When is practice testing most effective for improving the durability and efficiency of student learning?. *Educational Psychology Review*, *24*, 419-435.doi: 10.1007/s10648-012-9203-1

Rawson, K. A., Dunlosky, J., & Sciartelli, S. M. (2013). The power of successive relearning: Improving performance on course exams and long-term retention. *Educational Psychology Review*, *25*, 523-548. doi: 10.1007/s10648-013-9240-4

Reder, L.M. (1987). Selection strategies in question answering. *Cognitive Psychology, 19,* 90-138.doi: 10.1016/0010-0285(87)90005-3

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137,* 615-625. doi: 10.1037/a0013684

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*, 20-27. doi: 10.1016/j.tics.2010.09.003

Roediger, H.L. & Karpicke, J.D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249-255. doi: 10.1111/j.1467-9280.2006.01693.x

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140,* 1432-1463. doi: 10.1037/a0037559

Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence–accuracy relationship for eyewitness identification. *Law and Human Behavior*, *34*, 337-347. doi: 10.1007/s10979-009-9192-x

Schmalhofer, F., McDaniel, M. A., & Keefe, D. (2002). A unified model for predictive and bridging inferences. *Discourse Processes*, *33*, 105-132.doi: 10.1207/s15326950dp3302_01

Serra, M. J., & Dunlosky, J. (2010). Metacomprehension judgments reflect the belief that diagrams improve learning from text. *Memory, 18*, 698-711. doi: 10.1080/09658211.2010.506441

Shapiro, A., Sims-Knight, J., O'Rielly, G. V., Capaldo, P., Pedlow, T., Gordon, L., & Monteiro, K. (2017). Clickers can promote fact retention but impede conceptual understanding: The effect of the interaction between clicker use and pedagogy on learning. *Computers & Education*, *111*, 44-59.doi: 10.1016/j.compedu.2017.03.017

Siedlecka, M., Paulewicz, B., & Wierzchoń, M. (2016). But I was so sure! Metacognitive judgments are less accurate given prospectively than retrospectively. *Frontiers in psychology*, *7*, 218.doi: 10.3389/fpsyg.2016.00218

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 592. doi: 10.1037/0278-7393.4.6.592

Smith, S.M., Bjork, R.A., & Glenberg, A. (1978). Environmental context and human memory. *Memory & Cognition, 6,* 342-353.doi: 10.3758/bf03197465

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, *10*, 176-199. doi: 10.1177/1745691615569000

Stark-Wroblewski, K., Kreiner, D. S., Clause, C. B., Edelbaum, J., & Ziser, S. B. (2006). Does the generation effect apply to PowerPoint handouts? Psychology of Education, 43, 28-38. doi: 10.1037/e516682007-001

Storm, B. C., Friedman, M. C., Murayama, K., & Bjork, R. A. (2014). On the transfer of prior tests or study events to subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 115. doi:10.1037/a0034252

Szabo, A., & Hastings, N. (2000). Using IT in the undergraduate classroom: should we replace the blackboard with PowerPoint?. *Computers & education*, *35*, 175-187. doi: 10.1016/s0360-1315(00)00030-0

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences, 110*, 6313-6317. doi: 10.1073/pnas.1221764110

Szpunar, K. K., Jing, H. G., & Schacter, D. L. (2014). Overcoming overconfidence in learning from video-recorded lectures: Implications of interpolated testing for online education. *Journal of Applied Research in Memory and Cognition, 3*, 161-164. doi: 10.1016/j.jarmac.2014.02.001

Thorndike, E. L. (1914). Educational psychology, Vol 3: *Mental work and fatigue and individual differences and their causes*. New York, NY: Teachers College

Van Meter, P., Yokoi, L., & Pressley, M. (1994). College students' theory of note-taking derived from their perceptions of note-taking. *Journal of Educational Psychology*, *86*, 323-338. doi: 10.1037//0022-0663.86.3.323

Vaughn, K. E., Dunlosky, J., & Rawson, K. A. (2016). Effects of successive relearning on recall: Does relearning override the effects of initial learning criterion? *Memory & cognition*, *44*, 897-909. doi: 10.3758/s13421-016-0606-y.

Vaughn, K. E., & Rawson, K. A. (2011). Diagnosing criterion-level effects on memory: What aspects of memory are enhanced by repeated retrieval?. *Psychological Science*, *22*, 1127-1131.doi: 10.1016/j.jml.2014.04.004

Vekaria, P. C. (2011). *Lecture note-taking in postsecondary students with self-reported attention-deficit/hyperactivity disorder* (Doctoral dissertation, Columbia University).

Verkoeijen, P. P., Rikers, R. M., & Özsoy, B. (2008). Distributed rereading can hurt the spacing effect in text memory. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *22*, 685-695.doi: 10.1002/acp.1388

Weatherly, J. N., Grabe, M., & Arthur, E. I. (2003). Providing introductory psychology students access to lecture slides via Blackboard 5: A negative impact on performance. *Journal of Educational Technology Systems*, *31*, 463-474. doi: 10.2190/KRW7-QHFY-AY3M-FFJC

Weener, P. (1974). Note taking and student verbalization as instrumental learning activities. *Instructional Science*, *3,* 51-73.doi: 10.1007/bf00117026

Weinstein, Y., Gilmore, A. W., Szpunar, K. K., & McDermott, K. B. (2014). The role of test expectancy in the build-up of proactive interference in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 1039-1048. doi: 10.1037/a0036164

Weinstein, Y., Nunes, L. D., & Karpicke, J. D. (2016). On the placement of practice questions during study. *Journal of Experimental Psychology: Applied, 22*, 72-84. doi: 10.1037/xap0000071

Wollen, K. A., Cone, R. S., Margres, M. G., & Wollen, B. P. (1985). Computer programs to facilitate detailed analysis of how people study text passages. *Behavior Research Methods, Instruments, & Computers*, *17*, 371-378.doi: 10.3758/bf03200944

Williams, R. L., & Eggert, A. C. (2002). Notetaking in college classes: Student patterns and instructional strategies. *The Journal of General Education*, 173-199. doi:10.1353/jge.2003.0006

Wissman, K. T., & Rawson, K. A. (2015). Grain size of recall practice for lengthy text material: Fragile and mysterious effects on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 439-455. doi: 10.1037/xlm0000047

Wongkietkachorn, A., Prakoonsuksapan, J., & Wangsaturaka, D. (2014). What happens when teachers do not give students handouts?. *Medical teacher*, *36*, 789-793. doi: 10.3109/0142159X.2014.909921

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General, 145,* 918-933. doi: 10.1037/xge0000177

Accompanying Materials

Paper 1

Lecture habits survey

Part 1: Demographics. Please answer the questions below:

1) Your student ID:
2) Your year of study:
3) Your age:
4) Your subject:
5) Your gender:

Part 2:

Have you, at least once in the last semester annotated printed notes or made your own notes using either pen and paper or your computer, tablet or smartphone?

Y/N

Please indicate whether you take notes in lectures using pen and paper or if you take notes on your computer/tablet/smartphone:

- Pen and Paper (take them to questions related to pen and paper notes in part 4)
- Computer/tablet/smartphone (take them to questions related to computer/tablet/smartphone notes in part 4)
- I do not make notes

Part 3: please answer the following questions, giving as much detail as you wish, relating to the scenario below:

You are attending one of your weekly lectures in your subject and you know that you will have an end of module exam which may include material based on this lecture. Your lecturer has provided the slides to print ahead of time if you wish.

Describe what you do during a lecture like this one (for example, take notes, annotate the slides, test yourself on the material in the break?)

Even if they are not your personal habits, what would you are the best things to do in this lecture to ensure you remember as much as possible?

How do you try to remember the information given in a lecture for later use?

How do you use the lecture material (such as your own notes or slide printouts), if at all, to revise for the end of module test?

What do you think a tutor or lecturer can do to help you maximise the long-term retention of the information in a lecture? (For example, do you think they should provide a certain type of handout to accompany their lecture?)

Part 4: Please indicate whether these statements apply to you (where it says handwritten notes, this would be replaced with computer/tablet/smartphone notes if they checked that they take notes on these devices in part 2).

1) If lecturers put their slides on blackboard, I print them off and take them to the lectures
Y/N

2) I like the lecturers to provide handouts of their slides (or make these available for me to print).
Y/N

3) I always make handwritten notes, regardless of whether a copy of the slides is available
Y/N

4) I make less handwritten notes when a lecturer provides, or makes available, handout slides
Y/N

5) I think all lecturers should provide slide handouts Y/N

6) I believe that annotating slide handouts is a sufficient aid to allow me to remember the content of the lectures for later recall Y/N

7) I find the lecture more difficult when I do not have access to a slide handout Y/N

8) I share my handwritten lecture notes with my friends Y/N

9) I think my handwritten notes are good quality Y/N

10) I use my handwritten notes for memory prompts when revising Y/N

Part 5:

a) Please use tick boxes to indicate which of the methods you would use during a normal lecture:

(Please include everything, even factors you have already indicated above)

Make notes on plain paper (not the slide handout)

Make notes on a computer/smartphone/tablet in a separate document to the lecture presentation

Make notes on the slide handouts

Read accompanying literature (such as journal articles or book chapters)

Observe slide handouts without annotation

I observe lectures without writing or using any other method

Other (Please specify):

b) Please tick whichever boxes apply to your writing style during a lecture"

I paraphrase what the lecturer says into notes

I paraphrase the slides into my own notes

I try to write down everything the lecturer says

I try to copy down all the information on the slides

I write down salient words for later

I do not write anything during lectures

Other (Please specify)

Experiment 1 – Verbal Transcripts of Lecture Voiceover

Influenza

1) This is a presentation on Influenza, particularly the prevalence and pathology behind the disease.

2) Commonly referred to as flu, influenza is an infectious virus caused by the RNA family Orthomyxoviradae, which is transmitted though the air in coughs and sneezes. It is frequently confused with the common cold but is more serious and is the cause of approximately half a million deaths per year

3) Common symptoms of influenza are:

Fever - particularly extreme cold and shivering

Cough

Nasal congestion

Aching throat and limbs

Fatigue

Headache

Gastrointestinal symptoms – such as diarrhea and vomiting. These are particularly common in children

4) There are three types of influenza virus:

Influenza A – which is most common virus amongst humans and is responsible for the most severe disease form. It is primarily transmitted from wild aquatic mammals and a recent serotype, which means subviral category, was responsible for the outbreak of swine flu, in 2009.

Influenza B – Almost exclusively affects humans but is less common than influenza A. A degree of immunity can be acquired at a young age

Influenza C- is the least common type of virus. Which usually only mildly affects young children.

All have a similar overall structure and mechanism of infection.

5) The structure of the influenza virus. The particle is 80-120 nm long and is spherical in shape. All viruses are made up of a viral envelope containing two glycoproteins wrapped around a central core. Glycoproteins are involved in the interaction between cells. The centre of the virus consists of viral RNA genome and proteins to protect it, which you can see in the diagram.

6) Replication and spreading of the virus. Viruses only replicate in living cells and the influenza replication and infection takes place over a number of steps.

7) Firstly the virus binds to a host cell through haemogluttinin, a substance that causes red blood cells to clot but in this case, just acts as a transport means. A channel is made, by which endocytosis, engulfing of the surrounding cell constituents, occurs.

8) Once the virus is inside the cell, through a series of events caused by acidic conditions, the viral envelope is broken down and RNA is released into the cytoplasm

9) The core proteins in the cell and the viral RNA (vRNA) are then transported into the nucleus where transcribing occurs. This is the encoding and replicating of vRNA. Newly synthesized viral proteins are created, which will be discussed in more detail on the next slide. The original virus remains in the cell, often interfering with the messenger RNA involved in healthy cell replication.

10) The newly created viral proteins are packed with RNA in the cells to create Virons, which leave the nucleus and begin membrane protrusion, where the viruses adhere to the cell by the haemagluttinin, as seen when entering, and leave the cell by an exit, created by neuroamenodaze which, in this case, breaks down the haemagluttinin. After all the virons have been released, the host cell dies

11) Vaccinations can prevent the spread of influenza and are recommended in the UK for

Children

The Elderly

Asthma sufferers

Diabetics

Immunocomprimised people

Those with heart disease

12) There are no specific treatments for influenza. Paracetamol can alleviate symptoms but does not speed recovery. Young people are advised to avoid aspirin, as it can cause liver damage in combination with flu in this age group. As it is a viral infection, not a bacterial one, antibiotics will not have an effect.

Tuberculosis

1) This is a presentation on tuberculosis.

2) Tuberculosis is an infectious disease that most frequently affects the lungs and is caused by bacteria - usually mycobacterium tuberculosis in humans. Tuberculosis is fatal in approximately 50%of cases and is spread by airborne saliva particles such as coughing, sneezing and spitting. One in ten people who have the infection develop the active form of the disease.

3) The symptoms of tuberculosis are:

Chronic cough – with blood-tinged sputum

Fever

Weight loss

Night Sweats

And a tendency to fatigue easily

4) Tuberculosis is primarily caused by mycobacterium tuberculosis, or MBT, which is a small, aerobic bacteria with a high lipid content that divides very slowly, at the rate of approximately once per 16 to 20 hours. MBT can withstand weak disinfectants and can survive in a dry state for weeks

5) These are MBT under an electron microscope. They are approx 2um long and they do not have phospholipid outer membranes so are classified as gram positive bacteria. You can't see it in this diagram but gram positive refers to the fact that they are able to take up a gram stain used and turn purple. If this were a colour image you would see them with a purple tinge. However, if a gram stain is used, the high lipid content of the bacteria means that they do not stain very strongly so would more likely be a light pinky purple

6) MBT is part of a complex that contains 4 other TB causing bacterium:

M.Bovis – more commonly found in humans before pasturised milk

M.Africanum – not widespread, but still causes tb in some underdeveloped isolated communities in central Africa

M. Canetti – similarly, rare, most frequently found amongst Africans and occasionally, African immigrants

M. Microti – usually seen in immunodeficient people

7) The infection of tuberculosis begins when the bacterium reach the pulmonary alveoli, the small air sacs in the lungs highlighted here in the diagram.

8) They invade the endosomes, a membrane bound compartment inside the alveolar macrophages, which are the white blood cells in the alveolar spaces.

The primary site of the infection is the Gonn Focus and is the space between the upper and lower lobes, as pointed to here by the red arrow in this diagram. The infection then spreads around the lungs and further into the kidneys, brain and bone, which are the most common parts of the body to develop a secondary infection

9) Tb can be described as a granulomatous, inflammatory disease. The word "granulomatous" refers to the white blood cells, the macrophages, T-lymphocytes, B-lymphocytes and fibroblasts, which aggregate together to form granulomas with lymphocytes, which are other white blood cells, around the outside, which can be seen in the diagram here. Notice how the bacteria are taking up a purple stain thus are gram positive. The bacteria remain dormant in these granulomas and remain dormant in latent infection

10) A common feature of granulomas is the development of abnormal cell death called necrosis. An affected tissue in the lungs becomes widespread and appears as cheese like lumps called caseous necrosis, which can be seen in the diagram. As you can see, the Gonn Focus, which we discussed previously, is the area filled with the white, cheese like caseous necrosis. If untreated, pulmonary tuberculosis can lead to pneumonia

11) TB was highly common in the UK until the mid 20th century when the BCG vaccine was established. It is still common in Africa, where is affects 363 per 100,000 as opposed to the 15 per 100,000 affected in the UK. It is believed that the high incidence of HIV, causing immunosuppressant effects, contributes to the increase in TB, rates as well as less medical help available in more underdeveloped countries. Like other bacterial infections, TB continues to develop new strands that require constant research and continuous development of new drugs. However, in the UK the BCG vaccine appears to be maintaining the low incidences TB occurring.

Experiment 1 – Multiple Choice Tests

Multiple Choice Quiz – Influenza.

Instructions:

- Answer the 15 questions below, placing a tick in the box underneath your chosen answer. Please answer all of the questions even if you are unsure, there is no negative marking.
- Only select ONE answer for each question, if you select more than one answer, the question will be marked as wrong
- Underneath your chosen answer, alongside the row labelled “confidence”, please give a rating (as a percentage) of how confident you are that your chosen answer is correct.
- The participant with the highest combined score across all of their tests will win a £25 Amazon voucher. You will be informed if you are the winner by email once all the data has been collected.

Please read all the questions carefully, begin when you are ready:

1) What is the name of the RNA family that causes influenza?

	A	B	C	D	E
Answers:	Chrysoviridae	Birnaviridae	Orthomyxoviridae	Hypoviridae	Totiviridae
ıosen answer:					
Confidence:					

2) Which one of these is NOT a common symptom of influenza?

	A	B	C	D	E
Answers:	Blooded sputum	Vomiting	Diarrhoea	Night sweats	Headache
ıosen answer:					
Confidence:					

3) Which type of virus only affects young children?

	A	B	C	D	E
Answers:	Influenza A	Influenza B	Influenza C	Influenza D	Influenza E
Chosen answer:					
Confidence:					

4) **Which type of virus is responsible for the most severe form of influenza?**

	A	B	C	D	E
Answers:	Influenza A	Influenza B	Influenza C	Influenza D	Influenza E
Chosen answer:					
Confidence:					

5) What is the approximate size of an influenza virus particle?

	A	B	C	D	E
Answers:	100 picometres	100 nanometres	100 micrometres	100 millimetres	100 centimetres
Chosen answer:					
Confidence:					

6) What is the role of the glycoprotein in the core of the influenza virus structure?

	A	B	C	D	E
Answers:	Transport	Packaging	Interactions	Nutrients	Protection
Chosen answer:					
Confidence:					

7) How many glycoproteins are present in each virus structure?

	A	B	C	D	E
Answers:	2000	2	3	4	50

ıosen answer:					
Confidence:					

8) **Which word is used to describe the process when a virus engulfs and destroys the surrounding cell membranes?**

	A	B	C	D	E
Answers:	Endocytosis	Hypocytosis	Orthocytosis	Exocytosis	Oxocytosis
ıosen answer:					
Confidence:					

9) **Which of the following is a required condition for viral replication to occur?**

	A	B	C	D	E
Answers:	High humidity	Low humidity	Alkaline	Acidic	Warmth
ıosen answer:					
Confidence:					

10) **Which part of the virus breaks down to enable RNA to be released into the host cell cytoplasm?**

	A	B	C	D	E
Answers:	Nucleus	Vacuole	Ribosomes	Viral envelope	Viral strands
ıosen answer:					
Confidence:					

11) **What is the name given to the encoding and replicating of viral RNA?**

	A	B	C	D	E
Answers:	Transpiring	Transcribing	Translating	Transencoding	Transcripting
ıosen answer:					

Confidence:					

12) Which substance causes red blood cells to clot?

	A	B	C	D	E
Answers:	Haemoglobin	Viral RNA	Haemagluttinin	Clotting agent F	Phospholipids
Chosen answer:					
Confidence:					

13) What is the final outcome for a host cell invaded by viral RNA?

	A	B	C	D	E
Answers:	Becomes RNA	Cell Mutation	Cell Death	Returns to normal	Cell Replication
Chosen answer:					
Confidence:					

14) Which of the following is not considered "high risk"?

	A	B	C	D	E
Answers:	The elderly	Pregnant women	Diabetics	Asthmatics	Children.
Chosen answer:					
Confidence:					

15) Which of the following, in combination with influenza, is associated with liver damage?

	A	B	C	D	E
Answers:	Paracetamol	Penicillin	Insulin	Heparin	Aspirin
Chosen answer:					
Confidence:					

Multiple Choice Quiz – Influenza.

Instructions:

- Answer the 15 questions below, placing a tick in the box underneath your chosen answer. Please answer all of the questions even if you are unsure, there is no negative marking.

- Only select ONE answer for each question, if you select more than one answer, the question will be marked as wrong
- Underneath your chosen answer, alongside the row labelled "confidence", please give a rating (as a percentage) of how confident you are that your chosen answer is correct.
- The participant with the highest combined score across all of their tests will win a £25 Amazon voucher. You will be informed if you are the winner by email once all the data has been collected.

Please read all the questions carefully, begin when you are ready:

1) How is influenza transmitted between people?

	A	B	C	D	E
Answers:	Skin Contact	Particles in the air	Salivary contact	Through blood	Sharing needles
Chosen answer:					
Confidence:					

2) What is the basic structure of the influenza virus made of?

	A	B	C	D	E
Answers:	DNA	SNA	RNA	TNA	VNA
Chosen answer:					
Confidence:					

3) Which common ailment bears most similarity to influenza?

	A	B	C	D	E
Answers:	Diarrhoea	Indigestion	Cold	Headache	Earache
Chosen answer:					
Confidence:					

4) Which of the following is a common symptom of influenza in adults?

	A	B	C	D	E
Answers:	Vomiting	Diarrhoea	Chest pain	Shivering	Choking
ıosen answer:					
Confidence:					

5) Which influenza virus type was responsible for the outbreak of Swine flu in 2009?

	A	B	C	D	E
Answers:	Influenza B	Influenza C	Influenza C1	Influenza C2	Influenza A
ıosen answer:					
Confidence:					

6) Which influenza virus type is almost entirely exclusive to humans?

	A	B	C	D	E
Answers:	Influenza B	Influenza C	Influenza C1	Influenza C2	Influenza A
ıosen answer:					
Confidence:					

7) Which structural part of influenza A is different from B and C?

	A	B	C	D	E
Answers:	Nucleus	Shape	They have the same overall structure	Viral envelope	Internal proteins
ıosen answer:					
Confidence:					

8) How many glycoproteins are wrapped around the central core of the virus?

	A	B	C	D	E
Answers:	1	12	1000	2	5
Chosen answer:					
Confidence:					

9) What is the name of the substance which causes red blood cells to clot and acts as a transport means during the spread of the influenza virus?

	A	B	C	D	E
Answers:	Haemagluttinin	Aggregluttinin	Polygluttinin	Monogluttinin	Oxygluttinin
Chosen answer:					
Confidence:					

10) What condition is required for the activity of viral replication to occur?

	A	B	C	D	E
Answers:	High temperature	Alkaline	Low temperature	Acidic	Tepid conditions
Chosen answer:					
Confidence:					

11) What is the name of the viral proteins packed with genetic material which are created inside the host cell?

	A	B	C	D	E
Answers:	Virons	Monoglutimates	Polypeptide C	Oxidocarmamase	Monopeptide A
Chosen answer:					
Confidence:					

12) Which of the following people would be recommended for a flu vaccine?

	A	B	C	D	E

Answers:	18 year old student	36 year old many working in a hospital	29 year old woman with epilepsy	8 year old girl	52 year old dentist
ıosen answer:					
Confidence:					

13) Again, which of the following people would be recommended for a flu vaccine?

	A	B	C	D	E
Answers:	Heavily overweight 40 year old man	19 year old who smokes and drinks excessively	17 year old homeless girl	37 year old pregnant woman	54 year old taxi driver
ıosen answer:					
Confidence:					

14) What is a potential risk of combining aspirin with the influenza virus in teenagers?

	A	B	C	D	E
Answers:	Brain tumour	Liver damage	Kidney disease	Diabetes	Heart disease
ıosen answer:					
Confidence:					

15) What is the only cause of a disease which can be cured by antibiotics?

	A	B	C	D	E
Answers:	Muscular	Viral	Bacterial	Autoimmune	Lymphatic
ıosen answer:					
Confidence:					

Multiple Choice Quiz – Tuberculosis.

Instructions:

- Answer the 15 questions below, placing a tick in the box underneath your chosen answer. Please answer all of the questions even if you are unsure, there is no negative marking.

- Only select ONE answer for each question, if you select more than one answer, the question will be marked as wrong
- Underneath your chosen answer, alongside the row labelled "confidence", please give a rating (as a percentage) of how confident you are that your chosen answer is correct.
- The participant with the highest combined score across all of their tests will win a £25 Amazon voucher. You will be informed if you are the winner by email once all the data has been collected.

Please read all the questions carefully, begin when you are ready:

1) What percentage of people who have the tuberculosis infection develop the active disease?

	A	B	C	D	E
Answers:	10%	50%	1%	0.1%	20%
ıosen answer:					
Confidence:					

2) Which of the following is NOT a symptom of tuberculosis?

	A	B	C	D	E
Answers:	Fever	Night sweats	Blooded sputum	Diarrhoea	Weight loss
ıosen answer:					
Confidence:					

3) What is the full name for MBT, the bacteria responsible for the tuberculosis infection?

	A	B	C	D	E
Answers:	Mycobacterium	Myobaccilus	Mycobacterium	Mycobaccilus	Myobacterium

	tuberlus	tuberculosis	Tuberculosis	tuberlus	Tubercus
Chosen answer:					
Confidence:					

4) Which of the following would not affect the activity of MBT?

	A	B	C	D	E
Answers:	Moderately high temperatures	Moderately low temperatures	Weak disinfectants	High acidity	High alkalinity
Chosen answer:					
Confidence:					

5) How long is the average time taken for MBT to divide?

	A	B	C	D	E
Answers:	18 days	18 hours	18 minutes	18 seconds	18 nanoseconds
Chosen answer:					
Confidence:					

6) MBT are gram positive bacteria. What colour do they turn under a gram stain?

	A	B	C	D	E
Answers:	Pink	Orange	Green	Blue	Purple
Chosen answer:					
Confidence:					

7) Which of the following forms of tuberculosis is most common in immunodeficient people?

	A	B	C	D	E
Answers:	M. Canetti	M. Microti	M. Africanum	M. Bovis	M. Tuberculosis

ıosen answer:					
Confidence:					

8) Which of the following forms of tuberculosis is exclusive to isolated cultures is central Africa?

	A	B	C	D	E
Answers:	M. Canetti	M. Microti	M. Africanum	M. Bovis	M. Tuberculosis
ıosen answer:					
Confidence:					

9) What is the name of the primary site of infection?

	A	B	C	D	E
Answers:	Ghon Focus	Pulmonary alveoli	Lower left lobe	Upper right lobe	Purkinje fibres
ıosen answer:					
Confidence:					

10) Which of the following is NOT one of the white blood cells invaded by the bacteria?

	A	B	C	D	E
Answers:	Macrophage	T. Lymphocyte	B. Lymphocyte	C. Leukocyte	Fibroblast
ıosen answer:					
Confidence:					

11) Which of the following is NOT a common primary or secondary infection location?

	A	B	C	D	E
Answers:	Lungs	Kidneys	Intestines	Brain	Bone
ıosen answer:					

Confidence:					

12) What is the name of the congregated areas of abnormal cell death?

	A	B	C	D	E
Answers:	Coagulative necrosis	Gummatous necrosis	Fatty necrosis	Caseous necrosis	Liquefactive necrosis
Chosen answer:					
Confidence:					

13) What colour is this substance?

	A	B	C	D	E
Answers:	Black	Red	Clear	Green	Pale yellow
Chosen answer:					
Confidence:					

14) How many people, per 100,000, were affected by tuberculosis in the UK in 2010?

	A	B	C	D	E
Answers:	1.5	15	150	1500	15000
Chosen answer:					
Confidence:					

15) Why are people with HIV more susceptible to tuberculosis?

	A	B	C	D	E
Answers:	Depressed immune system	Lower lung capacity	Higher white blood cell content	Higher platelet content	Higher blood pressure
Chosen answer:					

Confidence:					

Multiple Choice Quiz – Tuberculosis.

Instructions:

- Answer the 15 questions below, placing a tick in the box underneath your chosen answer. Please answer all of the questions even if you are unsure, there is no negative marking.

- Only select ONE answer for each question, if you select more than one answer, the question will be marked as wrong
- Underneath your chosen answer, alongside the row labelled “confidence”, please give a rating (as a percentage) of how confident you are that your chosen answer is correct.
- The participant with the highest combined score across all of their tests will win a £25 Amazon voucher. You will be informed if you are the winner by email once all the data has been collected.

Please read all the questions carefully, begin when you are ready:

16) In what percentage of cases is tuberculosis fatal?

	A	B	C	D	E
Answers:	5%	15%	50%	65%	95%
Chosen answer:					
Confidence:					

17) Which of the following is a common symptom of tuberculosis?

	A	B	C	D	E
Answers:	Vomiting	Sweating	Headache	Earache	Chest pain
Chosen answer:					
Confidence:					

18) How large is a tuberculosis bacterium (MBT)?

	A	B	C	D	E
Answers:	2 nanometres	2 micrometres	2 millimetres	2 picometres	2 decimetres
Chosen answer:					
Confidence:					

19) Which of the following is in high content in the MBT?

	A	B	C	D	E
Answers:	Peptides	Sugars	Lipids	Monoglycerates	Glycerol
ıosen answer:					
Confidence:					

20) Which TB causing bacteria was commonly affecting humans before pasteurised milk?

	A	B	C	D	E
Answers:	M. Microti	M. Bovis	M. Canetti	M. Africanum	M. Tuberculosis
ıosen answer:					
Confidence:					

21) Which TB causing bacteria is usually seen in people with HIV?

	A	B	C	D	E
Answers:	M. Microti	M. Bovis	M. Canetti	M. Africanum	M. Tuberculosis
ıosen answer:					
Confidence:					

22) At which part of the lungs does the infection begin to invade surrounding cells?

	A	B	C	D	E
Answers:	Bronchi	Bronchioles	Lower lobes	Upper lobes	Alveoli
ıosen answer:					
Confidence:					

23) Which component of the blood is invaded by the bacteria?

	A	B	C	D	E
Answers:	Plasma	Platelets	White blood cells	Red blood cells	Any abnormal cell

Chosen answer:					
Confidence:					

24) What colour is caseous necrosis?

	A	B	C	D	E
Answers:	Red	Yellow	Black	Green	Pink
Chosen answer:					
Confidence:					

25) What is the cause of caseous necrosis?

	A	B	C	D	E
Answers:	White blood cell death	Red blood cell death	Bacterial cell death	Pulmonary tissue death	Any abnormal cell death
Chosen answer:					
Confidence:					

26) Which of the following is the most common secondary infection from tuberculosis?

	A	B	C	D	E
Answers:	Muscular	Blood stream	Bone	Lymphatic system	Nervous system
Chosen answer:					
Confidence:					

27) If untreated, what does tuberculosis lead to?

	A	B	C	D	E
Answers:	Pneumonia	Influenza	Whooping cough	Rubella	Polio
Chosen answer:					

Confidence:					

28) When was the BCG vaccination established?

	A	B	C	D	E
Answers:	17th century	18th century	19th century	20th century	21st century
ıosen answer:					
Confidence:					

29) Approximately how many people, per 100,000 were affected by tuberculosis in Africa in 2010?

	A	B	C	D	E
Answers:	4	40	400	4000	40,000
ıosen answer:					
Confidence:					

30) Approximately how many people, per 100,000 were affected by tuberculosis in Europe in 2010?

	A	B	C	D	E
Answers:	0.15	1.5	15	150	1500
ıosen answer:					
Confidence:					

Experiment 2 – Verbal Transcripts of Lecture Voiceover

Acid Rain

Slide 1

This is an introductory lecture about Acid Rain. How it occurs and its effect.

Slide 2

In the most simple terms acid rain is, as the name suggests, any form of precipitation in which the pH is less than 7, so is acidic. However, the pH of acid rain is usually no lower than 5.7.

Slide 3

Acid rain particularly affects the following groups:

Soil and plant life, particularly forests. This is because Some microbes within the soil are unable to tolerate changes to low pH and are killed. The deficiency in these microbes affects the plant life and affects sensitive species, such as the sugar maple tree.

Freshwater fish. Freshwater has a completely neutral pH and this is the water that these fish thrive in. When the pH is reduced, the more sensitive fish eggs will not hatch and in more extreme acidities, the fish will die.

Buildings made of limestone or marble. This is the most famous example of acid rain, as it is the most obvious. The effect will be explained over the next few slides.

Finally, whilst acid rain causes no direct damage to human health, the particulates responsible for the acid rain have an adverse effect.

Slide 4

So, what causes acid rain? The particles involved are:

Sulphur Dioxide

Nitrogen Dioxide

Hydroxyl Radicals

The first two particles are released from power plants during fuel combustion, whereas hydroxyl radicals are naturally occurring in the atmosphere.

Slide 5

In the first stage of the procedure, Sulphur dioxide reacts with the naturally occurring hydroxyl radical. A hydroxyl radical is a hydrogen molecule attached to an oxygen molecule. This radical is highly reactive so reacts quickly with other molecules:

$SO_2 + OH\cdot \rightarrow HOSO_2\cdot$

The compound which is produced, is unstable and reacts quickly with oxygen in the air to form sulphur trioxide and a hydroperoxyl radical

$HOSO_2\cdot + O_2 \rightarrow HO_2\cdot + SO_3$

In the presence of the naturally occurring water droplets in the air, the sulphur trioxide is converted to sulfuric acid

$SO_3 (g) + H_2O (l) \rightarrow H_2SO_4 (aq)$

The aq means that it is dissolved in water.

Slide 6

The process involving nitrogen dioxide is much more simple. It reacts with the naturally occurring hydroxyl radical discussed in the previous slide to form nitric acid

$NO_2 + OH\cdot \rightarrow HNO_3 (aq)$

Again, the aq means that it is dissolved in water.

Slide 7

Although acid rain has occurred throughout the presence of human life on earth, its prevalence significantly increased during the industrial revolution, as the production of fuel became increasingly important, thus the quantities of Sulphur Dioxide and Nitrogen Dioxide produced increased. Acid rain was first identified in the 17^{th} Century by John Evelyn who noticed a decline in marble statues over time.

The reason why acid rain causes damage to marble and limestone buildings and statues is that a key element is calcium carbonate ($CaCO_3$, which is an alkaline substance). The calcium carbonate reacts with the sulphuric acid in the acid rain to create a neutral substance, gypsum, which flakes away from the building or statue, along with harmless carbon dioxide and water:

$CaCO_3 (s) + H_2SO_4 (aq) \rightleftharpoons CaSO_4 (s) + CO_2 (g) + H_2O (l)$

Slide 8

The areas most affected by acid rain are:

- Eastern Europe

- The eastern third of USA

- Southwestern Canada

- Taiwan

- South Eastern coast of China

There are measures in place to attempt to reduce the acid rain in the environment. For example, the clean air agreement of 1991, and, many power stations now have manual methods of removing sulphur gases from their emissions. These measures are effective and the affects of this type of air pollution are reducing with each decade.

Ozone

Slide 1

This is an introductory lecture about the depletion of the ozone layer, how this occurs and its effects.

Slide 2

The **ozone layer** or **ozone shield** refers to a region of Earth's stratosphere that absorbs most of the sun's ultraviolet (UV) radiation. It contains high concentrations of ozone (the chemical O_3). However, although it is high compared to other parts of the atmosphere it is still only 10 parts per million of ozone in the ozone layer, compared to 0.3 parts per million in other parts of the atmosphere.

This Ozone layer absorbs between 97 and 99% of the suns ultraviolet light. This UV light is damaging to life forms, thus the ozone layer plays a vital role in maintaining plant and animal life. However, since 1970 the ozone layer has reduced in size by 4%.

Slide 3

Over the past 40 years, scientists have identified that the reason for this depletion is because of the increased levels of Cloroflurocarbons (CFCs) released into the atmosphere. These CFCs are found in common household items such as:

Refridgerators

Aerosols

Cleaning Solvents

Older Air-conditioning units

Blowing agents for foams

Slide 4

So how are these CFCs affecting the ozone layer. Firstly, you need to know the structure of the ozone layer itself. Oxygen exists in three forms in the atmosphere; O (atomic oxygen), O_2, the common gas and by far the most common form, and Ozone (O_3). In the presence of Ultraviolet light, the commonly occurring O_2 breaks down to form two oxygen atoms which then combine with separate O_2 molecules to form Ozone.

Slide 5:

However, Ozone is easily broken down from:

$3O_3 \rightarrow 2O_2$

Which is the most stable form of oxygen. This is an ongoing process within the Ozone layer and is necessary in order for the UV light to be absorbed, thus to prevent the radiation hitting earth (as discussed previously). However, this happens in equilibrium with the formation of Ozone discussed in the previous slide. The fact that the Ozone layer is not being maintained is due to the increased breakdown of Ozone to oxygen, which is due to the presence of these CFC molecules

Slide 6:

Choroflurocarbons are not highly reactive, thus are not broken down readily in the atmosphere before reaching the ozone layer. When they reach the layer, they are broken down by electromagnetic radiation:

$CFCl_3$ + electromagnetic radiation$\rightarrow Cl\cdot + \cdot CFCl_2$

A two step process then begins in which each Cl molecule (which in this form is called a radical) breaks down two Ozone molecules.

Firstly, The Cl radical combines with ozone to form Chlorine Monoxide and Oxygen:

$Cl\cdot + O_3 \rightarrow ClO + O_2$

This Chlorine monoxide then destroys a second ozone molecule to recreate the original CL radical:

$ClO + O_3 \rightarrow Cl\cdot + 2O_2$

The fact that the chlorine radical remains means that it can continue to break down these ozone molecules indefinitely through this process.

Although the above example used chlorine, bromine, fluorine and iodine radicals can cause the same damage when part of the original hydrocarbon.

Slide 7

The Ozone depletion is particularly apparent over Antarctica, where Ozone levels are 33% lower than in 1975. This has now started to expand and low levels of Ozone are being detected in Australia, New Zealand, Chile, Argentina and South Africa.

Although there is an increase in awareness of the depletion of the Ozone layer, people frequently confuse the effects with those of Global Warming. Whilst both cause change within the ecosystem on earth, the depletion of the Ozone layer should not have a direct effect on the earth's temperature, just the presence of UltraViolet light.

Slide 8

In light of increased understanding as to the causes of Ozone depletion, the United States National Academy of Sciences has recommended the elimination of CFCs from aerosol spray cans and, as of the 1980s, companies tend to use alternative substances in aerosols instead of CFCs. However, a suitable replacement has not been found for refrigerants or cleaning circuit boards. In some counries, CFC-free refrigerants have been trialed although these alternatives are mostly greenhouse gases which in turn cause environmental inequalities. Because of these CFCs, the process of disposal of used refridgerators in becoming increasingly controlled and increasing policies are in place to prevent incorrect disposal.

Experiment 2 – Cued-recall tests

Acid Rain – Questionnaire A

Please read the questions below and try to answer them to the best of your ability.

- All questions should be answered in no more than a couple of sentences maximum
- Please write something for every question. Please do not leave blanks
- Please avoid writing "I don't know". Make an attempt at writing something based on your memory of the presentation
- Underneath each question is a box with a space in which to write a percentage. For each question, please indicate how confident that your answer is correct as a percentage (0%-100% with 100% being entirely sure your answer is correct.)
- If you have any questions, please alert the experimenter and they will come and answer your questions to the best of their ability.

Thank you and Good Luck!

Please turn the page over when you are ready

1) What is the lowest pH of acid rain which is usually found?

Answer = 5.7

2) Why is soil affected by acid rain?

Answer = Because microbes within the soil cannot tolerate low pH and are killed.

3) Name a building material which is affected by acid rain:

Answer = Limestone or Marble

4) What is the name of the naturally occurring radical which is involved in the production of acid rain?

Answer = Hydroxyl Radical

5) In the presence if naturally occurring water droplets in the air, what is sulphur trioxide converted to?

Answer = Sulphuric acid

6) Complete the formula: $SO_2 + OH\cdot \rightarrow$ ________

Answer = $HOSO_2$

7) What does (aq) mean in $H_2SO_{4\ (aq)}$?

Answer = Dissolved in water

8) When Nitrogen Dioxide reacts with the hydroxyl radical, what acid is formed?

Answer = Nitric Acid

9) What historical event caused the increase in the production of acid rain?

Answer = The Industrial Revolution

10) What is the common name given to the solid, neutral substance which flakes away from buildings containing calcium carbonate in the presence of acid rain?

Answer = gypsum

11) Which constituent of the acid rain reacts with calcium carbonate to cause damage to some buildings?

Answer = sulphuric acid

12) Eastern Europe and Taiwan are particularly affected by acid rain. Name another area in the world that was mentioned in the presentation and particularly affected

Answer = Southwestern Canada, The eastern third of USA or the South Eastern coast of China.

Acid Rain – Questionnaire B

Please read the questions below and try to answer them to the best of your ability.

- All questions should be answered in no more than a couple of sentences maximum
- Please write something for every question. Please do not leave blanks
- Please avoid writing “I don’t know”. Make an attempt at writing something based on your memory of the presentation
- Underneath each question is a box with a space in which to write a percentage. For each question, please indicate how confident that your answer is correct as a percentage (0%-100% with 100% being entirely sure your answer is correct.)
- If you have any questions, please alert the experimenter and they will come and answer your questions to the best of their ability.

Thank you and Good Luck!

Please turn the page over when you are ready

1) What is the pH of an acidic substance?

Answer = Less than 7

2) Which animals are the most directly affected by acid rain?

Answer = Freshwater fish

3) What industrial process causes Sulphur Dioxide and Nitrogen Dioxide to be released into the atmosphere?

Answer = Fuel Combustion

4) Complete the sentence "Sulphur Dioxide reacts with the naturally occurring ______________ radical to form an unstable product"

Answer = Hydroxyl

5) What is SO_3 known as?

Answer = Sulphur Trioxide

6) What is the common name given to H_2SO_4?

Answer = Sulphuric Acid

7) Complete the equation: $NO_2 + OH\cdot \rightarrow$ ________

Answer = HNO_3

8) When was acid rain first identified?

Answer = 17th Century

9) How was acid rain first identified?

Answer = A decline in marble statues over time

10) What is the key element in marble and limestone buildings which reacts with sulphuric acid?

Answer = Calcium Carbonate (Or $CaCO_3$)

11) Why does it react with sulphuric acid?

Answer = It is acidic

12) What decade was the Clean Air Agreement released?

Answer = 1990s

Ozone Depletion – Questionnaire A

Please read the questions below and try to answer them to the best of your ability.

- All questions should be answered in no more than a couple of sentences maximum
- Please write something for every question. Please do not leave blanks
- Please avoid writing "I don't know". Make an attempt at writing something based on your memory of the presentation
- Underneath each question is a box with a space in which to write a percentage. For each question, please indicate how confident that your answer is correct as a percentage (0%-100% with 100% being entirely sure your answer is correct.)
- If you have any questions, please alert the experimenter and they will come and answer your questions to the best of their ability.

Thank you and Good Luck!

Please turn the page over when you are ready

1) What is another name for the Ozone layer?

Answer = The Ozone Shield

2) What percentage of the sun's UV light is absorbed by the Ozone layer?

Answer = 97-99%

3) Other than Aerosols, name a household item that contains Chloroflurocarbons.

Answer = Refrigerators, Cleaning Solvents, air conditioning units or blowing agents for foams.

4) In which form does the oxygen molecule most commonly, and most stably, exist?

Answer = O_2

5) What is this stable form broken down to in the presence of UV light?

Answer = Oxygen atoms (O)

6) Complete the equation: $2O_3 \rightarrow$ ______

Answer = $3O_2$

7) Why are CFCs not broken down in the atmosphere before reaching the Ozone layer?

Answer = They are not highly reactive

8) What is the name given to this type of $Cl\cdot$ molecule which breaks down Ozone molecules

Answer = A radical

9) What happens to this Cl· after it has converted Ozone to Oxygen gas?

Answer = it continues to break down more Ozone molecules

10) Where is the Ozone depletion most apparent?

Answer = Antarctica

11) Which phenomenon is frequently confused with the Ozone depletion?

Answer = Global Warming.

12) In what decade did companies begin to remove CFCs from their aerosol propellants?

Answer = 1980s

Please read the questions below and try to answer them to the best of your ability.

- All questions should be answered in no more than a couple of sentences maximum
- Please write something for every question. Please do not leave blanks
- Please avoid writing “I don’t know”. Make an attempt at writing something based on your memory of the presentation
- Underneath each question is a box with a space in which to write a percentage. For each question, please indicate how confident that your answer is correct as a percentage (0%-100% with 100% being entirely sure your answer is correct.)
- If you have any questions, please alert the experimenter and they will come and answer your questions to the best of their ability.

Thank you and Good Luck!

Please turn the page over when you are ready

1) What is the main purpose of the Ozone layer?

Answer = To absorb UV radiation

2) How many parts per million of Ozone are present in the Ozone layer?

Answer = 10 parts per million

3) How many parts per million of Ozone are present in other parts of the earth's atmosphere?

Answer = 0.3 parts per million

4) Approximately how much, as a percentage, has the Ozone layer reduced in size since 1970?

Answer = 4%

5) What is the chemical symbol for Ozone?

Answer = O_3

6) What has to be present for Oxygen gas molecules to break into Oxygen atoms?

Answer = Ultraviolet light

7) What do the Oxygen atoms combine with to form Ozone?

Answer = Oxygen gas molecules

8) Complete the equation:

$CFCl_3$ + electromagnetic radiation→ ____+ $\cdot CFCl_2$

Answer = Cl·

9) What is the name given to the chemical ClO?

Answer = Chlorine Monoxide

10) How many Ozone molecules are destroyed by a chlorine radical in each two-step process?

Answer = Two

11) What happens to the Chlorine radical after the process is complete?

Answer = It remains to break down more Ozone particles

12) Other than Antarctica, name a country which is detecting low levels of Ozone:

Answer: Australia, New Zealand, Chile, Argentina, South Africa

Paper 2 – Verbal Transcript of Lecture Material

1. This is a presentation on MRSA. I would like you to observe it like a normal lecture. However, please do not write on the sheets in front of you until there is a break to do so. You will be asked to summarise the key points of the previous three slides. These are in the slides in bold font. You will be given a minute to do this. Please write the points in full sentences so that you could use them for later revision.

2. MRSA stands for Methicillin Resistant Staphylococcus Aureus. Methicillin is the name of an antibiotic and Staphylococcus Aureus is the bacteria which is resistant to methicillin.
3. Staph A most commonly colonises in the nostrils, and is found there. It can exist harmlessly as the natural flora but it is problematic and more prevalent in high density populations or where patients have invasive devices, where there in access for the bacteria to enter. It is most commonly found in hospitals.
4. MRSA can remain asymptomatic for many years, residing naturally in the body flora. However, when it takes form as an infection, the most common symptom is the appearance of small red bumps which develop into pus filled boils when untreated. This can progress to the appearance of fever and other rashes which are similar to those seen in meningitis,
5. As said before, it is asymptomatic until the immune system becomes weakened and the bacteria is allowed to spread. When it does so, this occurs quickly, often spreading over 24/48 hours and can cause widespread infections and fevers, examples of which are Toxic Shock syndrome and Necrotising pneumonia.
6. MRSA is easy to detect. A simple nostril swap using a cotton wool bud which is then smeared on a petri dish can be used to isolate Staph A. This picture shows what a staph A culture looks like on a petri dish. However, as we have said, this does not necessarily mean this person will become ill with MRSA, but they are a risk if their immune system weakens or if they are exposed to people with weakened immune systems.
7. Staph A is a gram positive bacteria – gram positive means that when it is exposed to a gram stain it will turn purple – this is because it has a lipid outer layer which allows the gram stain to enter the cell. However this is only visible under a microscope.
   Staph A is also anaerobic so it is capable of surviving without oxygen, and it is round in

   appearance and tends to form clusters like in this picture.
8. 20% of the entire population are carriers of Staph A however this does not affect them as it resides in the natural flora of their nasal passages
9. Stap A is not exclusive to humans, other species listed above have been found with the bacteria. Like with humans, in a weakened immune system, it can cause problems such as bumblefoot in chicken and mastitis in cows.
10. So Staph A is the bacteria, let's move on to what methicillin is. Methicillin is an antibiotic from the penicillin class, which is used to treat infections from gram positive bacteria, which, as we discussed earlier are a type of bacteria that has a positive gram stain result. We know Staph A is a gram positive bacteria, so it looks like it should be affected by the actions of this antibiotic.
11. SO how does it work? It prevents the bacteria from developing cell walls and without cell walls they cannot replicate or function as bacteria. It does this at the molecular level. Polymer chains like this form links between the chains to create the cell wall, as you can see in the bottom picture. However, the methicillin inhibits these links so the wall cannot be constructed and the bacteria cannot grow or spread.
12. Methicillin is not affected by penicillinase. This is a bacterial enzyme which has been produced to counterbalance the effects of the antibiotics. We will discuss this and how it works in more detail over the next few slides.
13. So penicillinase is also known as beta lactamase and it is an enzyme which is produced by the bacteria as a resistance to the antibiotics in the penicillin class.
14. Most antibiotics have a molecule called beta lactam in their structure. This is a four atom ring (as shown in the diagram) which is targeted by penicillinase, hence why it is otherwise known as beta lactamase.
15. This beta lactam molecule is broken down by penicillinase to deactivate the antibiotic and it does this through hydolysis. This is where chemical bonds which are part of the beta lactam molecule are broken down into two separate molecules and water molecules. This is a complex process and you do not need to know any more detail except hydrolysis breaks down larger molecules to smaller molecules plus water molecules and that some

antibiotics, if they have this beta lactam molecule, are destroyed by penicillinase in this way but this does not occur in the case of methicillin.

16. So how is Staph A resistant to methicillin if it is not because of the production of the enzyme penicillinase to break down the antibiotic?
    It is actually at the genetic level that staph A is resistant to methicillin. This is a complex process. The gene responsible for this resistance is called Mec A and it is found in all genes within cells of staph A, the diagram shows this.
17. Mec A is called a mobile genetic element. This means it can move within all of the genetic material of an organism, so it is not restricted to a specific location. This Mec A codes for the resistance of Methicillin.
18. As we saw on the diagram, Mec A is integrated into every Staph A chromosome but can move freely within that. It codes for resistance. This is done by creating a replacement unit which binds to the bacterial DNA in place of the antibiotic DNA. Therefore there is no space for the antibiotic to bind to the DNA thus it cannot break down the polymer chains and inhibit the cell wall growth we discussed previously thus the bacteria continue to replicate and antibiotic, in this case, the methicillin, is ineffective.
19. There are many different subgroups and strands of MRSA in the world. Within the UK, the two most commonly seen strands are MRSA 15 and 16, which each have their own infection patterns.
20. eMRSA 15 originated in kettering, but is also now found in South East Asia so has spread as people either visited the UK and contracted the bacteria or people from the UK have visited and passed the bacteria to residents. It is lesser known than eMRSA16.
21. eMRSA 16 is identical to one of the most prevalent strands in the USA, probably also due to cross contamination. It is more common throughout the world than eMRSA 15. it is also associated with the onset of toxic shock syndrome is extreme cases.
22. Now I will briefly discuss the history, prevalence and incidence of MRSA within the UK. It was licenced in 1595 to treat Staph A, which had become resistant to other penicillin antibiotics due to the production of the beta lactamase/penicillinase enzyme secreted by the bacteria that we discussed previously. Although, as we know, methicillin was not affected by penicillinase, by 1961, MRSA had already been identified in 1961.
23. So on to incidences. MRSA increased in prevalence until 2006, where it spiked with 1,652 deaths caused directly from MRSA, a 51% increase from 13 years previously, showing how quickly the resistance to the antibiotic developed.
24. In terms of the prevalence, whilst the number of people with Staph A in their natural flora is relatively constant from what we know, the rate of infection varies around the world and is affected by many factors such as rate of disease, health vigilance and climate conditions. Current estimates are that in developed countries, Romania has the highest infection rate per person with Staph A in their system and Iceland has the lowest rate of infection.
25. In terms of treatment, there are some newer antibiotics, such as Daptomycin, which are effective at killing the staph A virus. However, as we have seen, bacteria will eventually become resistant to this new antibiotic, so these must be used sparingly to reduce exposure to the bacteria as the overuse of anitbiotics has contributed to the development of resistance of previous antibiotics by bacterial strands. New pathways to reduction of bacterial are being sought.
26. Is is generally considered better to prevent the infection than to attempt to treat it. However, when 20% of the population have the bacteria residing naturally in their system, this is more difficult as they are a threat when exposed to people with weakened immune systems or if their immune systems weakens.
27. To reduce cases of infection, screening programmes have been suggested in which nasal swabs are taken in at-risk areas such as hospitals. This would involve taking nasal swabs and testing them for the presence of Staph A as shown earlier/ This would alert the

carriers to the presence of the bacteria and prevent them from contacting at-risk populations.

28. Equally, surface sanitizaton has shown a reduction in bacterial levels and pure alcohol has been found to be more effective than any other sanitizer.
29. Many MRSA antibiotics are either in phase 2 (large scale testing of healthy controls) or phase 3 (testing on sufferers to test for an effect) of clinical drug trials to attempt to find a more effective, stronger antibiotic which could cure MRSA.
30. Equally, some natural remedies have been suggested to reduce the spread of the infection. Maggots clean the dead tissue to reduce infection and semi toxic mushrooms and some cannibinoids excrete antibiotics excrete antibiotics and have been effective at preventing further spread of bacteria.
31. In 2004, researchers here at The University of Southampton demonstrated that the spread of MRSA is reduced in the presents of copper alloys at room temperature as copper has antimicrobal properties. However, this has not yet entered trials so it is unknown whether it would have an effect on human subjects with MRSA.

Paper 2 – Test Questions and Correct Answers plus Rationales for Incorrect Options

What does MRSA stand for in its entirety?

A) Methicillin Resistant Staphyloccocus Aureus

Where is Staph A most commonly found?

1) Nostrils – Correct answer

2) Throat – Same passageway
3) Intestines – Also a commonly affected bacterial passage
4) Brain – less commonly affected by bacteria
5) Nerves – Unaffected by bacterial, no blood

In which of the following scenarios is MRSA the biggest threat?

1) Hospitals – Correct answer
2) Prison – Also mentioned in the presentation
3) School – An area of high population but not permanently residing and generally healthy
4) House – Area of no population
5) Outdoors – Area of least threat due to lack of enclosure and low population

Which is the primary symptom of MRSA?

1) Red Bumps – Correct answer
2) Yellow Pimples – Mentioned in the presentation as a secondary symptom
3) Black spots – Still related to the presence of a rash, a symptom of other bacterial infections
4) Nausea – A symptom of other bacterial infections
5) Back pain – Related to nervous problems and rare in infections

MRSA is normally asymptomatic, what happens to a person to cause the onset of symptoms?

A) Weakened immune system

How can MRSA be detected?

1) Nostril Swab – Correct answer
2) Throat swab – Same system (respiratory)
3) Blood test – Would find some bacteria in blood
4) MRI – bacteria cannot be seen on an MRI but still acknowledges that it is something that would not be externally present
5) Gait test – Testing walking would not indicate a presence of bacteria, which cannot be seen externally

What colour is a positive gram stain?

1) Purple – correct answer
2) Pink – similar
3) Blue – similar but would indicate negative result
4) Yellow – Never applicable to a gram stain
5) Black and orange – implausible

What percentage of people have Staph A residing in their system?

1) 20% - Correct answer
2) 30% - Closest incorrect answer
3) 5%
4) 50%
5) 80% - most inaccurate incorrect answer

Which of the following species of animal has been found to carry Staph A?

1) Cow – Correct answer
2) Sheep – Also a farm animal
3) Goat – Less common farm animal
4) Camel – Found in different location

5) Whale – not a land animal

Which class of Antibiotic is methicillin in?

A) Penicillin Class

Which part of the cell synthesis is inhibited by methicillin?

1) Cell wall – Correct answer
2) Nucleus – Common cell component also involved in cell replication
3) Cytoplasm – Just the solution in which the cells are held, not involved in replication
4) Vacuole – Not part of a bacterial cell
5) Molecule – Not a cell constituent

What action does penicillinase have on methicillin?

A) No action

What is another name for Beta Lactamase?

1) Penicillinase – Correct answer
2) Methicillinase – Incorporates the specific bacterial name
3) Alpha lactamase – Incorporates the "-ase" suffix given to all enzymes
4) Enzyme 659 – acknowledges that it is an enzyme
5) Streptococcus B – A bacteria

What molecule is targeted by penicillinase?

1) Beta Lactam – Correct answer
2) Alpha Lactam – Second part is correct
3) Cell wall – Is the structure targeted but not an individual molecule
4) Nucleus – not part of the cell targeted
5) The enzyme – non sensical

By what process does Beta Lactamase deactivate the Beta Lactam molecule?

1) Hydrolysis – Correct answer
2) Metabolism – A process of breakdown of molecules but incorrect
3) Catabolism – A process of molecule change, but growth instead of breakdown
4) Aerobic decomposition – Not a biological term but the word "decomposition" acknowledges breakdown
5) Photosynthesis – A term only applying to plants and no human cells

What is Staph A's resistance to methicillin based on?

1) A gene - Correct answer
2) A molecule – A particle constituent of a gene
3) An atom – A particle constituent but smaller and less plausible
4) An electron – Smaller and less plausible than an atom
5) An enzyme – Discussed in presentation as not affected by the enzyme

What is the name of the gene that codes for methicillin resistance in Staph A DNA?

1) Mec A – Correct answer
2) Mec M – Contains "Mec" and M for methicillin acknowledging Methicillin resistance
3) Gene A – Acknowledging "A"
4) Lysine – A DNA component
5) Meth A – Not a gene

How does Mec A prevent the antibiotic (Methicillin) from binding with the Staph A?

A) Create a replica unit

eMRSA 16 is one of the most common strands of MRSA in the UK, what is the name of the other strand?

A) eMRSA 15

Other than the UK, where is eMRSA 15 also found?

A) Asia

Which disease can occur following an onset of eMRSA 16?

1) Toxic Shock Syndrome – Correct Answer
2) Necrotising Pneumonia – a symptom of other strands of MRSA
3) Tuberculosis – A disease caused by bacteria
4) Influenza – A disease caused by virus
5) Vomiting – Not a disease

Which year was Methicillin launched for treatment?

1) 1959 – Correct answer
2) 1961 – nearest incorrect date
3) 1952
4) 1971
5) 1929 – furthest date from correct answer

Approximately how many deaths were caused by MRSA in its most prevalent year, 2006?

1) 1500 – Correct answer
2) 1000 – nearest incorrect answer
3) 100
4) 15
5) 10,000 – furthest incorrect answer

Romania has the highest rate of infection from MRSA, which European country has the lowest rate of infection?

A) Iceland

Why are newer antibiotics used sparingly now in medicine?

A) To avoid resistance strands from forming in the bacteria

Screening programmes are recommended but not used within the UK – what is the method of this screening?

1) Nostril Swab – Correct answer
2) Throat swab – Same system (respiratory)
3) Blood test – Would find some bacteria in blood
4) MRI – bacteria cannot be seen on an MRI but still acknowledges that it is something that would not be externally present
5) Gait test – Testing walking would not indicate a presence of bacteria, which cannot be seen externally

What sanitizer is recommended as the most effective at reducing Staph A?

1) Pure Alcohol – Correct Answer
2) Antiseptic handwash – Antiseptic is antibacterial so would reduce bacteria
3) Solvent cleaner – Generally effective as a cleaner but not as effective as those above
4) Germicide – kills germs not necessarily bacteria
5) Iodine – Non alcohol so the opposite of the most effective

Which of the following natural products inhibits the spread of MRSA?

1) Maggots – Correct answer
2) Leaches – An insect known for healing properties
3) Royal Jelly – An animal product with healing properties
4) Calendula – An antifungal plant but not antibacterial
5) Hemp – no known medicinal properties

What phase/s of clinical trials are potential new antibiotics to treat MRSA currently in?

A) 2&3

In 2004, the University of Southampton demonstrated that which metal inhibited the spread of MRSA?

1) Copper – Correct answer
2) Iron – Also a metal with known medical properties
3) Gold – a metal but with no medicinal properties
4) Brass – Not a true metal but a metal alloy
5) Potassium – Not a metal in any form

Paper 3

Experiment 1 – Verbal Transcript

IVF

SLIDE 1 –Intro 1

This is a presentation on IVF. Today I will be talking about what it is, its history, the procedure and the outcomes.

SLIDE 2 –Intro 2

IVF stands of in-vitro fertilization. It is the process of using sperm to fertilise an egg outside of the female body, before returning the fertilised egg to the female uterus with the intention of establishing a successful pregnancy.

SLIDE 3 – Intro 3

Because of this, IVF is considered as a type of assisted reproductive technology. There are other types of assisted reproductive technology, some occurring outside of the body, such as ICSI, or intracytoplasmic sperm injection, and some occurring inside the body, such as or IUI, intrauterine insemination. However, for this lecture, we will just discuss simple cases of IVF.

SLIDE 4 – Definition 1

The term “in vitro” is from the latin word meaning glass. This is because the first biological experiments into this procedure were carried out in glass test tubes. This also has led to the term “test tube baby” as a phrase for a child who is conceived and born following this procedure.

SLIDE 5 - Definition 2

In the scientific community, the term “in vitro” is now used to refer to any biological procedure which occurs outside of a living organism. This helps to distinguish it from “in vivo”, which refers to procedures where the tissue remains inside the living organism.

SLIDE 6 - Definition 3

These days, the IVF procedure usually takes place in shallow petri dishes, which are made of a non-organic plastic material. However, some IVF methods are performed on organic material, but are still considered “in vitro”.

SLIDE 7 – Who has IVF? 1

IVF is most commonly used to overcome female infertility in heterosexual couples. A common reasons for this infertility is blocked fallopian tubes, preventing the egg from descending prior to fertilization. It can also be used in occasions in which a male has reduced sperm quality or a low sperm count.

SLIDE 8 - Who has IVF? 2

However, it is becoming increasingly common that IVF is used for same sex couples to conceive a child. In the case of a female same sex couple, this is often through reciprocal IVF, in which one partner provides the egg, which is fertilized and placed in the uterus of the other female. In the case of a male same sex couple, this is when a fertilized egg is placed inside a surrogate.

SLIDE 9 - Who has IVF? 3

Due to improved medical technology, IVF can also be used in the case in which certain genetic characteristics are required or desired. This can range from a couple choosing IVF to select the gender of their child to a couple using IVF due to one of the parents being a carrier of a genetic disorder. The fertilized eggs can then be selected for implantation based on their characteristics.

SLIDE 10 – Procedure: Egg retrieval 1

Prior to the retrieval of eggs from the female, drugs are provided to induce ovarian hyperstimulation. This is when the female is injected with gonadotropins, a hormone group which stimulates the body to produce an excess of follicle stimulating hormone (FSH).

SLIDE 11 - Procedure: Egg retrieval 2

FSH, or follicle stimulating hormone is naturally synthesized and secreted in women. It encourages the early growth and recruitment of ova, or eggs, within the ovaries. An excess of FSH caused by hyperstimulation will encourage the growth of more ova, thus more eggs will be produced in that menstrual cycle.

SLIDE 12 - Procedure: Egg retrieval 3

After a scan has confirmed that there are sufficiently mature ova present in ovaries, a thin needle is inserted and guided via ultrasound to the overies where the ova can be aspirated alongside their surrounding follicular fluid. It is common to remove between ten and 30 eggs in any single aspiration, and the procedure takes between 20 and 40 minutes. It can be done under either local or general anaesthetic.

SLIDE 13 – Procedure: Preparation and incubation 1

After the ova and sperm have been successfully transferred to the laboratory, the eggs are then stripped of surrounding cells and prepared for fertilization. A selection process may occur in which the most healthy eggs, most likely to sustain a successful pregnancy will be separated, and the rest destroyed. In the meantime, the sperm is prepared from the male semen sample through a process called sperm washing

SLIDE 14 - Procedure: Preparation and incubation 2

The ova and sperm are then incubated together at a ratio of approximately 75,000:1 in a culture in a petri dish in order for fertilization to take place. Nowadays, this is commonly for 1-4 hours, although it used to be for between 16 and 24 hours, but this was found to be less effective in a review from 2013.

SLIDE 15 - Procedure: Preparation and incubation 3

Laboratories have developed grading methods to judge embryo quality in order to optimise pregnancy rates. Embryo quality is measured mostly using microscopy and biomarkers such as RNA and proteins are observed, and examined for damage. However, if the IVF has occurred to avoid genetic disorders, as we discussed previously, a priority is given to the screening and selection of embryos which show the lowest risk of inheriting the genetic disorders carried by the parents.

SLIDE 16 – Procedure: Transfer of Embryos 1

The number of embryos transferred in this phase depends on a number of factors such as the age of the female, the number of unsuccessful IVD cycles that the donors have tried previously and the country in which the IVF is taking place, due to varying laws. However, to avoid multiple pregnancies, it is rare that more than two eggs are transferred

SLIDE 17 - Procedure: Transfer of Embryos 2

Embryo transfer occurs between 2 and 6 days following egg retrieval. This is to allow for specific screening which cannot be done until the embryo is more developed. The longer that the embryo is kept in vitro prior to transfer, the increased likelihood of live birth. However, the number of embryos available will be reduced so the risk has to be weighed on an individual basis.

SLIDE 16 - Procedure: Transfer of Embryos 3

Once again, ultrasound guidance is used to precisely place the embryos within the uterine cavity. The picture below shows two embryos following transfer into the uterus. The female then must wait between ten days and two weeks to test for pregnancy following transfer.

SLIDE 19 – Success rates 1

The success rate of IVF is defined as the number of live births per 100 rounds of IVF. It is important that this is distinguished from the pregnancy rate, as many factors affect how many pregnancies end in miscarriage, the most common of which is age.

SLIDE 20 – Success rates 2

The success rate of IVF in cases where the female is younger than 25 is 40.7 live births per 100 rounds. This is considered optimal success. The levels of success reduce with each two-year age increment and by the age of 43, the success rate is down to 3.9 live births per 100 rounds of IVF.

SLIDE 21 – Success rates 3

Apart from age, there are other predictors of IVF success. Some of which are biological, such as semen quality provided by the male donor, thickness of uterine lining and progesterone levels in the female, and some of which are related to lifestyle such as smoking, high BMI and high caffeine intake.

SLIDE 22 – Risks and Complications 1

A major complication of IVF is the risk of multiple births as multiple embryos are often transferred. Multiple births carry an increased risk of pregnancy loss, prematurity, neonatal morbidity and potential long-term damage. However, many people consider this to be a small risk as the chances of having healthy twins through IVF are high.

SLIDE 23 - Risks and Complications 2

With any medical procedure, there are risks involved to the egg provider and/or retriever. There is a small risk of bleeding, infection and damage to the surrounding structures such as the bowel, bladder and large intestines as well as the effects of the anaesthetic, whether local or general. These carry standard risks such as abdominal pain, nausea, vomiting and very rarely more serious side effects up to and including death.

SLIDE 24 – Risks and Complications 3

Regardless of the pregnancy result, one of the biggest risk factors considered by couples is the emotional stress caused by the treatment and potential negative outcome. Whilst social support has been found to have a relieving effect, the potential negative test result following a stressful procedure has been associated with increased depression rates in women.

SLIDE 25 – History and Incidence of IVF 1

The procedure of IVF was developed by Patrick Steptoe and Robert Edwards in the early 1970s and was based on early laparoscopy work completed by the two of them in Cambridge in the decade previously. Robert Edwards won the 2010 nobel prize in medicine for his work and Steptoe, already dead at the time, was awarded it posthumously.

SLIDE 26 - History and Incidence of IVF 2

The first baby successfully born through IVF was Louise Brown, who was born on 25th July 1978 at Oldham General Hospital, Manchester, UK. She was conceived on her parents' first attempt at the procedure following nine years of infertility

SLIDE 27 - History and Incidence of IVF 3

With the addition of egg donation, IVF can be used to help women conceive who are past their reproductive years. The oldest woman to conceive and give birth with IVF treatment was Romanian born Adriana Illescu, who was 66 at the time of the birth of her daughter. Both the sperm and egg used to conceive her daughter were donated anonymously.

SLIDE 28 – IVF around the world 1

In the USA, IVF is readily available at a cost averaging $12,400. IVF is not available with any public healthcare policy. The USA is considered to be one of the most costly countries for IVF around the world

SLIDE 29 - IVF around the world 2

In the UK, IVF can be available on the NHS and, in theory, every woman is eligible for at least one free cycle of IVF if they can demonstrate between 18 months and 3 years of fertility depending on area. However, there is no blanket treatment plan across the UK and the number of free cycles, the necessary duration and age of mother at the time of treatment vary from location to location.

SLIDE 30 - IVF around the world 3

As of 2012, every UN recognised country across the world allowed IVF in some form except for Costa Rica, which, until 2015, remained the only country with complete ban on IVF. However, on 10 September 2015, President Luis Guilliermo Solis legalised IVF. However, it is still being heavily contested by opposers in the country's constitutional court.

Experiment 1 – Test Questions and Answers

** Alternatives are ranked most – least plausible (1 – 5)

What does IVF stand for in its entirety?

A) In vitro fertilization

Which part of IVF takes place outside of the human body?

1) Egg fertilization
2) Egg extraction
3) Egg retraction

4) Egg creation
5) Egg retroaction

Which of the following is another type of assisted reproductive technology which occurs outside of the body?

1) ICSI
2) IUI
3) IDSI
4) IEI
5) IMEI

What is the translation of the latin phrase “in vitro”?

A) In Glass

What is the latin term used by medics to describe any procedure that takes place in the body?

1) In vivo
2) In utero
3) In vitro
4) En vitro
5) Il viro

Where does MRSA most commonly take place in modern day?

1) Petri dish
2) Non organic cylindrical beaker
3) Organic cylindrical beaker
4) Glass test tube
5) Glass cylindrical beaker

Which of the following is the most common reason for a female to need IVF?

1) Blocked fallopian tubes
2) Blocked ovarian ducts
3) Low quality eggs
4) Insufficient duodenal lining
5) Blocked seminal ducts

What is the name of the procedure in which a female donates her eggs to her female partner to carry her offspring

A) Reciprocal IVF

Which of the following is NOT a typical reason for IVF to be chosen for treatment?

1) To avoid having a child with down’s syndrome
2) To avoid having a child with sickle cell anaemia
3) To avoid having a child with Cystic fibrosis
4) To avoid having a male child
5) To avoid having a female child

What is the name of the hormone group which is injected into the female to cause ovarian hyperstimulation to begin the egg retrieval process?

1) Gonadotrophins
2) Follicle Stimulating Hormone
3) Prostaglandin
4) Progesterone
5) Adrenaline

What is the name of the hormone encourages the growth of more ova?

A) FSH (Follicle stimulating hormone)

Approximately how many eggs are taken from the ovaries in a single aspiration?

1) 20
2) 2
3) 200
4) 2000
5) 20,000

Whereabouts are the sperm and ova prepared prior to incubation?

A) A laboratory

In what ratio are sperm and ovum incubated together?

1) 75,000:1
2) 7500:1
3) 75:1
4) 750,000:1
5) 7,500,000:1

Which of the following is an example of a biomarker used to judge embryo quality?

1) RNA
2) DNA
3) Protons
4) Neutrons
5) Tissue type

Which of the following is a factor that would affect how many embryos were transferred back into the female?

1) Number of previous unsuccessful IVF attempts
2) Age of father
3) Family history of multiple births
4) Number of previous children
5) Age of embryo

What is the maximum number of days between egg retrieval and egg transfer when the embryos are not frozen?

1) 6 days
2) 5 days
3) 8 days
4) 15 days
5) 30 days

How long does a female need to wait after embryo placement to take a pregnancy test?

A) Two weeks

Fill in the missing word: “Success of IVF is determined by the number of live births per ____

rounds of IVF”?

1) 100
2) 1000
3) 10,000
4) 100,000
5) 1,000,000

What is the approximate success rate of IVF in women over 43?

A) 4

Which of the following is NOT a predictor of IVF success?

1) Weight of male
2) Age of female
3) Weight of female
4) Caffeine intake
5) Semen quality from sperm donor

What is the risk associated with transferring multiple embryos during IVF?

A) Multiple births

Which of the following is a risk associated with the medical procedure of IVF?

1) Uterine infection
2) Bladder infection
3) Kidney infection
4) Intestinal infection
5) Stomach infection

What is considered to be the biggest factor for consideration before embarking on IVF treatment?

B) Emotional stress involved

What year did Patrick Steptoe and Robert Edwards win the Nobel Prize for medicine?

1) 2010
2) 2000
3) 1990
4) 1980
5) 1970

What year was the first successful IVF baby born?

A) 1978

How old was the oldest woman to use IVF to conceive a child?

1) 66
2) 68

3) 72
4) 56
5) 46

Approximately how much does a round of IVF cost in the USA?

1) $12,000
2) $10,000
3) $16,000
4) $6,000
5) $1,200

How many rounds is the miminim the NHS will fund in cases of infertility in the UK?

1) One
2) Two
3) Three
4) Four
5) Five

Which country legalised IVF in 2015?

1) Costa Rica
2) Nicaragua
3) Colombia
4) Saudi Arabia
5) Australia