



# UNIVERSITY OF TWENTE.

Faculty of Behavioural,  
Management & Social sciences (BMS)

## THE EFFECTS OF QUIZZING IN RECORDED LECTURES ON TEST-ANXIETY AND DELAYED LEARNING OUTCOMES

**Felicia Elskamp**

M.Sc. Thesis Educational Science and Technology

February 2020

---

**First Supervisor:**

Dr. H. van der Meij

**Second Supervisor:**

Dr. A.M. van Dijk

Department of Instructional Technology  
Faculty of Behavioural, Management,  
and Social Sciences  
University of Twente  
P.O. Box 217  
7500 AE Enschede  
The Netherlands

---

### **Abstract**

There is an increasing pressure for secondary education to employ active learning strategies that focus on students' individual learning needs. Hence the flipped classroom, in which students prepare at home using recorded lectures, has gained widespread attention. However, effectively processing a recorded lecture can be problematic. Quizzing can be used to tackle this problem by providing re-exposure to content, fostering active processing, and preventing students from overestimating themselves. However, quizzing might be anxiety provoking, which is associated with a decrease in academic achievement. A controlled, pre- posttest experiment within a real classroom setting generated new insights in the effects of quizzing in recorded lectures on delayed learning outcomes among pre-university students, with test anxiety as a mediating variable.

Three main conclusions can be derived from the empirical study: 1) quizzing does not improve delayed learning outcomes when factors such as external motivation, the frequency and the level of practice are the same for all students; 2) quizzing neither reduces nor increases test anxiety; 3) a high-quality lecture, either interpolated by quiz items or short summaries, can be used to enhance higher-order thinking. Results imply that re-exposure to content is effective when targeting the same content as the exam. Surprisingly, the quality of the lecture seemed to overrule the quizzing effects. This study adds high value to existing research on quizzing and recorded lectures since not only the effects of quizzing were investigated in a more controlled classroom setting, but also the effects on test anxiety were incorporated.

*Keywords.* [quizzing, test anxiety , delayed learning outcomes, recorded lecture, pre-university]

**I. Table of contents**

Abstract .....	2
I. Table of contents .....	3
II. List of Tables .....	5
III. List of Figures .....	5
IV. Acknowledgements .....	6
1. Introduction .....	7
2. Theoretical framework .....	9
2.1. Quizzing to enhance learning .....	9
2.2. Quizzing in recorded lectures .....	11
2.3. Test anxiety and learning .....	13
2.4. Test anxiety in relation to quizzing .....	15
2.5. Measuring learning outcomes.....	16
3. Research questions and Hypotheses .....	17
4. Method .....	18
4.1. Participants & Design.....	18
4.2. Instruments & Data analysis.....	19
4.3. Procedure.....	22
5. Results .....	25
5.1. Distribution of demographics .....	25
5.2. The effect of quizzing on video engagement.....	25
5.3. The effect of quizzing on learning outcomes .....	26
5.4. The effect of quizzing on confidence levels .....	28
5.5. The effect of quizzing on test anxiety .....	29
5.6. The mediating effect of test-anxiety on learning outcomes .....	30
6. Discussion .....	31
6.1. The effect of quizzing on video engagement.....	31
6.2. The effects of quizzing on learning outcomes.....	32

6.3. The effect of quizzing on test anxiety .....	36
6.4. Implications .....	37
6.5. Limitations.....	38
6.6. Future research .....	38
7. Conclusion.....	39
References .....	40
Appendix A – Test anxiety survey .....	45
A.1. Trait anxiety and demographics data.....	45
A.2. State anxiety survey.....	46
Appendix B – Pre domain knowledge test with self-reported confidence levels .....	47
Appendix C – Post domain knowledge test with self-reported confidence levels .....	49
Appendix D – Summaries used in the recorded lecture .....	51
Appendix E – Quiz items used in the recorded lecture .....	52
Appendix F – Logdata collected during the recorded lecture .....	54

## II. List of Tables

Table 1: Overview of the procedure per session.....	24
Table 2: Distribution of demographics among the three research conditions.....	25
Table 3: Distribution of males and females among the three research conditions .....	25
Table 4: Differences between video engagement measures in all three research conditions.....	26
Table 5: Differences in scores on the pre and post domain knowledge tests.....	26
Table 6: Differences between scores on low and high-level items.....	27
Table 7: Differences between scores on quizzed and non-quizzed items.....	28
Table 8: Differences in calibration accuracy in all three research conditions .....	29
Table 9: Differences in calibration bias in all three research conditions .....	29
Table 10: State anxiety during the pre- and post-test in all three research conditions .....	30

## III. List of Figures

Figure 1. Overview of the research design.....	19
Figure 2. Screenshots of the website used as the intervention.....	21
Figure 3. Mean test scores in the pre- and post-test for all research conditions.....	27

#### **IV. Acknowledgements**

I would like to express my very great appreciation to my supervisor, Dr. Hans van der Meij, for his valuable and challenging feedback throughout the entire process. By sharing his expertise he motivated me to put my best foot forward. I would also like to thank my second reader, Dr. Alieke van Dijk, who helped me to put the finishing touches on this project. My special thanks to Frank van den Belt and Gorgias Meijer, who allowed me to confiscate three of their valuable chemistry lectures to carry out the experiment. Moreover, the time and effort Frank put into helping me with creating the video, the quiz, and the test questions is highly appreciated. I am also very grateful for the assistance given by Henri Elskamp and Jeroen Waterink once my programming skills let me down. Without you, I would not have been able to develop the online lesson needed for my envisioned research design. Finally, I want to thank my family, friends and roommates for providing me with coffee and pep talks when I desperately needed it. A special thanks for Sierd, who was there for me throughout the entire rollercoaster ride. The help and support of all of you resulted in the thesis that now lies in front of you.

## 1. Introduction

There is an increasing pressure for secondary education to employ active learning strategies that focus on students' individual learning needs. As a result, the flipped classroom has gained widespread attention over the past years. In flipped classrooms, students prepare at home using recorded lectures and time in class is spent on deliberate practice (e.g. O'Flaherty & Phillips, 2015; Suo & Hou, 2017). This approach allows students to process material at their own pace and ask for personalized help during classroom activities (O'Flaherty & Phillips, 2015). However, research shows that effectively processing a recorded lecture can be problematic, because students tend to passively listen (Chi, 2009; O'Flaherty & Phillips, 2015) or overestimate themselves (Dunlosky & Rawson, 2012).

Research suggests that quizzing (i.e. the ungraded testing of educational content) can be used to overcome these obstacles. For example, Mayer et al. (2009) showed that real-time quizzing in class fosters active processing and improves students' scores on summative exams. Likewise, recent studies showed that quizzing improves students' processing of educational content by stimulating the use of effective learning strategies (García-Rodicio, 2015; Nguyen & McDaniel, 2014; Shapiro et al., 2017). In addition, Szpunar, Jing and Schacter (2014) stated that quizzing improves learning outcomes by helping students judge their performance and therefore prevent them from overestimating themselves. Altogether, this suggests that quizzing in recorded lectures can help students to effectively process the lecture and thereby improve their learning outcomes. However, it is unclear whether quizzing improves learning because of re-exposure to the same content, or because of the actual testing of knowledge. The current study, therefore, aims to investigate not only if but also in what way quizzing in recorded lectures can improve students' learning outcomes.

Moreover, there are ambiguities regarding the effects of quizzing on students' test anxiety, which is alarming because test anxiety is associated with a decrease in academic achievement (e.g. Ashcraft, 2002; Batchelor, 2015; Cassady, 2004). Some argue that quizzing can reduce test anxiety (Nyroos, Schéle & Wiklund-Hörnqvist, 2016) or has no effects (Khanna, 2015), whereas others state that quizzing is anxiety provoking (Crooks, 1988; Putwain, 2008). Therefore, when quizzing is implemented to increase learning outcomes, it is essential to consider possible opposed effects caused by test anxiety.

Nowadays, teachers start to recognize the added value of quizzing as a strategy to improve learning outcomes. However, they might not be aware of this strategy's boundary conditions and hence implement it ineffectively. For example, quizzing effects might not be significant when teachers implement quizzing without providing corresponding feedback (García-Rodicio, 2015). Moreover, problems arise when teachers implement quizzing without taking into account the difficulties students might experience because of test anxiety (Nguyen & McDaniel, 2014). Therefore, clearly outlining the effects of quizzing is of crucial importance to stimulate teachers to implement quizzing effectively. As

opposed to other recent studies, this study not only investigates the testing effect induced by quizzing, but a direct comparison to the effects of re-exposure to educational content is made. Deeper insights into these effects of quizzing will add to the available information about the use of quizzing in educational contexts. Moreover, there is a need for a study on the effects of quizzing on students' test anxiety, since this physiological condition is associated with a decrease in academic performance.

Altogether, this study aims to investigate the effects of quizzing in recorded lectures on delayed learning outcomes among pre-university students, with test anxiety as a mediating variable. This will be done by investigating the effects of quizzing using different versions of a recorded lecture in a pretest-posttest design. The majority of studies on the topic of quizzing are conducted in real classroom practices and these kinds of observational studies are afflicted by an omitted variables problem (Bruns, 2017). In this case, it means that it is not clear whether learning outcomes increased because of quizzing or because of, for example, emphasis on to-be-learned material, higher student motivation, or the amount of practice. A controlled experiment can more clearly isolate the effects of quizzing. Therefore, the current study was a controlled experiment within a real classroom setting, investigating whether the positive effects of quizzing found in literature also persist for real classroom practices in which some factors, such as external motivation, the amount, and the level of practice, are kept constant.



## 2. Theoretical framework

### 2.1. Quizzing to enhance learning

Quizzing can be defined as low-stakes testing of educational content (Dunlosky, Rawson, Marsh, Nathan & Willingham, 2013; McDaniel et al., 2011). In other words, quizzing is not used to assess performance, but to improve learning (e.g. Fiorella & Mayer, 2015; Nguyen & McDaniel, 2014). Research provides three explanations on how quizzing can improve learning: the re-exposure effect, active construction of knowledge, and improved metacognitive skills. The last two can be categorized under the testing effect.

#### Quizzing and the re-exposure effect

One of the possible explanations for the effectiveness of quizzing is the re-exposure effect. In this case, quiz items act as indicators of key concepts to help students recognize essential material of the video (Nevid & Mahon, 2009). These indicators can then be used to (re-)watch parts of the preceding video segments that are related to the quiz (Kovacs, 2016). However, opinions differ on whether the re-exposure effect induced by quizzing is beneficial or detrimental to learning.

On the one hand, re-exposure caused by quizzing is assumed to support learning by increasing the amount of information in memory and strengthening associations (Mayer, 1983). A study by Roelle, Roelle and Berthold (2018) supports this line of reasoning as it showed that quiz items which directed students' attention to a larger amount of the lesson content were more effective than quiz items targeting specific parts. Moreover, Mayer (1983) states that re-exposure not only affects *how much* is learned but also *what* is learned. According to him, re-exposure helps students to 1) focus on the main concepts of the provided information; 2) reorganise this information by relating key ideas to another and to existing knowledge; and 3) create a coherent whole by putting this information in their own words.

On the other hand, quizzing might restrict students' re-exposure to parts of the material that are targeted by the quiz, neglecting other important information that might be part of the summative assessment (Nguyen & McDaniel, 2014). A study by Kovacs (2016) showed that many students, instead of watching the entire video first, jump to the quiz to see what it is about and use that to navigate to parts of the video they believe are most important. This type of selective attention can harm learning because students might miss out on keys ideas needed to create a coherent whole. Multiple studies confirm this argumentation by showing that the learning effects of quizzed items do not persist for untargeted information (e.g. Nguyen & McDaniel, 2014; Shapiro, 2009).

To conclude, re-exposure caused by quizzing can improve recall by helping students to create a coherent whole of the presented material. However, students possibly focus on the quizzed material only and miss out on other essential information. Additionally, a study by McDaniel, Agarwal, Huelser, McDermott and Roediger (2011) showed that exposure per se (repeatedly presenting target content without the use of quizzing) can improve learning outcomes, but this effect is reinforced by adding quiz

items. Similarly, García-Rodicio (2015) showed that students who have to actively answer quiz questions outperform students who may look at the same question without the need of answering it. This indicates that quizzing, besides the re-exposure effect, induces another effect that influences students' learning outcomes: the testing effect.

### **Quizzing and the testing effect**

Another possible, widely documented explanation for the effectiveness of quizzing is the testing effect. The testing effect implies that students better remember material on which they have been tested than material that is merely restudied (e.g. Fiorella & Mayer, 2015; McDaniel et al., 2011). For example, McDaniel et al. (2011) found that eighth-grade science students who were quizzed a day before their final exam scored higher on the exam than students who were not quizzed. In this case, quiz items act as motivators to retrieve information from long-term memory. There are two prevailing explanations for the widely documented testing effect induced by quizzing.

First, quizzing stimulates active engagement (e.g. Mayer et al., 2009; Nguyen & McDaniel, 2014), which fosters a deeper understanding of the material (Shapiro et al., 2017). According to the SOI model (Fiorella & Mayer, 2015), students must *select* relevant material, mentally *organize* it, and then *integrate* it with prior knowledge to achieve meaningful learning. Quizzing is proven to be an effective learning strategy to support this process of generative learning (Dunlosky et al., 2013; Fiorella & Mayer, 2015; García-Rodicio, 2015). As described by García-Rodicio (2015), a quiz item requires students to choose the correct answer, which stimulates them to actively organize and integrate the information. Dunlosky et al. (2013) described this generative learning process more extensively: when students attempt to *select* target information needed to answer a quiz item, related information in their long-term memory is also activated and coded along with the target information. As a result, when students *integrate* the target information with prior knowledge, multiple pathways to the target and related information are created (Dunlosky et al., 2013). In other words, retrieving information from long-term memory to answer a quiz item helps students to mentally organize that information such that later retrieval becomes easier. This can be seen as active construction of knowledge. As opposed to short summaries, which can be neglected by the students, quiz items demand students to actively construct their knowledge (García-Rodicio, 2015). Therefore, it was expected that students who were presented with quiz items throughout a recorded lecture would have higher learning outcomes compared to students who were given short summaries instead.

Second, quizzing improves students' metacognition by helping them judge what they know and not know about the presented material (McDaniel et al., 2011; Szpunar et al., 2014). When providing students with feedback on quiz items, this effect can even be reinforced (García-Rodicio, 2015; McDaniel et al., 2011). Improved metacognition is expected to enhance learning, because by having a clear view of what they know and where they lack knowledge students can select more effective study strategies (Fiorella & Mayer, 2015; McDaniel et al., 2011). Moreover, if students are aware of a lack of

understanding, they can allocate additional cognitive resources to effectively process the provided feedback and adjust their understanding of the topic (García-Rodicio, 2015). Whereas students who receive short summaries instead of quizzes might overestimate their understanding of the topic (i.e. overconfidence) and will therefore not allocate additional resources to effectively process the summary content. Besides, accurately predicting their mastery of a topic might give students a feeling of control which in turn reduces test anxiety (Bledsoe & Baskin, 2014). This effect is discussed in more detail in section 2.4. To confirm that quizzing indeed improves metacognition, confidence (i.e. a dimension of metacognition) was measured in the current study by self-reported confidence levels during the pre- and post- domain knowledge tests. Using these confidence levels, the calibration accuracy (i.e. the absolute difference between expected and actual performance) and calibration bias (i.e. a measure of over- or underestimating performance) (Huff & Nietfeld, 2009) were calculated. It was expected that students who were presented with quiz items throughout a recorded lecture would more accurately predict their performance compared to students who were given short summaries instead.

In conclusion, quiz items can be used to foster re-studying of the material and/or stimulate active construction of knowledge and effective metacognition. To assess the degree to which these effects influence learning outcomes, three conditions were included in the current study. Students in the two test conditions were obligated to answer quiz items presented throughout the lecture, whereas students in the control condition were given short summaries containing information similar to that of the quiz. It was expected that students in the quizzing conditions would score higher on the post domain knowledge test compared to students in the control condition because: 1) quizzing stimulates active engagement and prevents students from neglecting the recap information (i.e. the quiz/summary) and 2) quizzing positively influences students' confidence levels, a dimension of metacognition that can improve students' performance on the post domain knowledge test. Additionally, students in one of the test conditions were allowed to re-watch the recorded lecture before answering quiz items, which presumably stimulates re-study of the material. In the other test condition, students were not allowed to look back at the recorded lecture before answering the quiz items, making active construction of knowledge essential. Based on the literature cited, it was expected that students who had to actively construct knowledge would score higher on the post-test compared to students in the other conditions.

## **2.2. Quizzing in recorded lectures**

Several things need to be considered when adding quiz items to recorded lectures to increase learning outcomes. First of all, the majority of research shows that quizzing is more effective when quiz items are supported by direct feedback (e.g. Agarwal, Karpicke, Kang, Roediger & McDermott, 2008; McDaniel, Anderson, Derbish & Morissette, 2007; Shapiro, 2009). For example, a study by Nguyen and McDaniel (2014) showed that no testing effect was found for quizzes that were not supported by elaborate feedback explaining which answer was correct and why. A possible reason for this effect is

that when provided immediately, feedback can be used to check one's understanding of the lecture material (i.e. metacognition) (García-Rodicio, 2015) and to correct ones misconceptions while the lecture material is still fresh (Fiorella & Mayer, 2015; Shapiro, 2009). However, some researchers suggest that anxiety is increased when students encounter failure (and thus negative feedback) during testing (Wise, Plake, Eastman, Boettcher, & Lukin, 1986). Fortunately, others showed that direct feedback reduced test anxiety for the majority of students (Attali & Powers, 2009; Dibattista & Gosse, 2006). Also, students indicated open questions *without* feedback as very stressful (Attali & Powers, 2006) which should therefore be avoided.

Secondly, the placement of the quiz items within the video should be considered. Quizzing is most useful after initial exposure to the lesson (Mayer, 2015) because this allows students to retrieve essential content (McDaniel et al., 2011). This does not necessarily mean that quiz items should be placed at the end of the lecture, placing them throughout the lecture might even be more effective (Szpunar et al., 2014). According to Glass (2009), quizzing is only effective when the interval between first encounter (the lecture) and second encounter (the quiz) with the study material is not too long, such that the initial representation of the information is still available and can be selected from memory. Therefore, quiz items in the current study were placed after the video segments in which essential information for answering that question was presented.

Thirdly, opinions differ on whether the quiz should be similar to the final exam. On the one hand, some argue that quiz items should closely match the final exam because students restrict their learning to the material shown in the quiz (Fiorella & Mayer, 2015; Roelle et al., 2018). In line with this argumentation, Shapiro (2009) states that the benefits of quizzing do not persist for information that is not addressed by one of the quiz items. On the other hand, many teachers do not want to use quiz items that are identical to questions in the final exam (McDaniel et al., 2007), because students would then be able to pass the exam by memorizing the correct answers rather than deeply understanding the material (Thomas, Weywadt, Anderson, Martinez-Papponi & McDaniel, 2018). Fortunately, research showed that quizzing can also enhance summative test performance when a concept is quizzed in one context and tested in another (Glass, 2009; McDaniel, Thomas, Agarwal, McDermott & Roediger, 2013). In the study of McDaniel et al. (2013) for example, the concept of 'competition for resources' was quizzed in a context of foxes and raccoons competing for pheasant. In the subsequent exam, the students' understanding of the same concept of competition was assessed in a different context, namely that of groups of pandas competing for bamboo. Altogether, the *context* might vary, but the quiz should address the same *concepts* as the exam in order to be effective.

Finally, the difference between low-level and high-level quiz items should be acknowledged. Whereas low-level questions simply ask students to retrieve essential information, high-level questions require students to go beyond the provided information (Roelle et al., 2018). High-level questions are expected to be more effective because they stimulate higher cognitive processing. This results in more coherent and accurate mental models (Roelle et al., 2018), allowing students to apply new knowledge

in more flexible ways (Thomas et al. 2018). However, some studies showed that low-level questions are more effective (Bing, 1982; Roelle et al., 2018), possibly because they can direct students to a bigger part of the lesson material (Roelle et al., 2018). So, it can be said that the effects of low- and high-level quizzing on exam performance are disputable. Some studies, therefore, included both low- and high-level questions when investigating the effects of quizzing. Thomas et al. (2018) showed that quizzing improved summative test scores regardless of the level of quiz items. In other words, factual quiz items not only improved performance on factual exam questions, but also on application exam questions. This is auspicious because when the level of quiz and exam questions can be varied, rote memorization of quiz answers will no longer be sufficient for students to score well on the summative exam (McDaniel et al., 2013).

In conclusion, quizzing is most effective when 1) supported by direct feedback; 2) placed throughout the recorded lecture, and; 3) addressing the same concepts as the summative exam. In the current study, quiz items were implemented in the recorded lecture accordingly. The effects of the level of quiz items are disputable and should be further investigated. The current study, therefore, implemented both low- and high-level quiz questions based on the first four levels Blooms taxonomy. Low-level questions included *remembering* and *understanding* (i.e. knowledge in a similar situation), whereas high-level question focused on *applying* (i.e. knowledge in a new situation), and *analysing* (i.e. knowledge of elements and their relations) (Krathwohl, 2002). Besides investigating the effects of quizzing in recorded lectures on students' learning outcomes, this study aims to explore how this effect is mediated by test anxiety. The following sections focus on the causes and effects of test anxiety and how this relates to quizzing.

### 2.3. Test anxiety and learning

Anxiety can be defined as “a state of apprehension, tension, or uneasiness that occurs in *anticipation* of internal or external danger” (Cummings, 1995, as cited in Bledsoe & Baskin, 2014, p. 33). The type of anxiety of interest for the current study is test anxiety, which is caused by concerns about one's test performance (Cassady, 2004; Covington & Omelich, 1987) and is widely associated with a decrease in academic achievement (e.g. Ashcraft, 2002; Batchelor, 2015; Cassady, 2004). Two types of test anxiety can be distinguished: trait test anxiety and state test anxiety.

Trait anxiety can be defined as anxiety that is experienced in any evaluative situation (Hong & Karstensson, 2002) and develops over time due to multiple causes found in the home and school environment. For example, high expectations and critical reactions of teachers and parents can lead to more anxious children who strive for approval by avoiding failure rather than approaching success (Wigfield & Eccles, 1989; Zeidner, 1998). Moreover, repeated failure can create a fixed mindset in which children believe they lack ability that cannot be improved, making them anticipate on failure and feel anxious, rather than being open to learning from mistakes (Bledsoe & Baskin, 2014; Wigfield & Eccles, 1989). In addition, children who believe they are equally or better skilled than peers feel less

anxious compared to children who believe the opposite (Lohbeck, Nitkowski & Petermann, 2016; Wigfield & Eccles, 1989).

State anxiety can be defined as anxiety that is only experienced in specific situations (Hong & Karstensson, 2002; Wigfield & Eccles, 1989), for example when making a, or studying for, a test. A possible cause for state anxiety is too complex tasks which can make the student feel out of control (Trevino & Webster, 1992). Multiple studies showed that a loss of control can increase state anxiety (Bledsoe & Baskin, 2014; Trevino & Webster, 1992). Besides the complexity of tasks, other factors that might cause a loss of control are time limits (Aydin, 2010; Wigfield & Eccles, 1989) and unstructured assignments (Wigfield & Eccles, 1989). Unstructured assignments make it more difficult for students to understand what is asked from them, which increases test anxiety (Wigfield & Eccles, 1989). Therefore, the video content and quiz items of the current study were divided among several manageable segments.

To conclude, test-anxiety can be measured in terms of trait and state anxiety. Since trait anxiety is slowly developed over a long time, a significant decrease in trait anxiety would most likely not be achieved within the scope of this study. Therefore, the focus of the current study was on state anxiety. State anxiety was measured during the pre- as well as the post domain knowledge test to investigate the effects of quizzing on state test anxiety (as of now simply referred to as test anxiety). Test anxiety was included in this study because, as mentioned before, it is associated with a decrease in academic achievement. The next section describes the effects of test anxiety on academic achievement in more detail.

### **The effects of test anxiety on academic achievement**

Besides physical effects like stomach ache and shortness of breath (Batchelor, 2015), test anxiety is associated with a decrease in academic achievement (e.g. Ashcraft, 2002; Batchelor, 2015; Cassady, 2004). Multiple effects of test anxiety on academic achievement can be found in literature.

First of all, the most well-known effect of test anxiety is anxiety blockage, which means that during an assessment, students are unable to retrieve previously learned information from long-term memory (Cassady, 2004; Covington & Omelich, 1987). According to Naveh-Benjamin, McKeachie and Lin (1987), students' worries about their abilities interfere with effective retrieval of information, causing the blockage. In addition, Covington and Omelich (1987) state that the initial study effort, either high or low, does not determine the degree of this interference. So, test anxiety can lead to poor academic achievement, even for students who prepared well for the test.

Secondly, test anxiety not only affects students' abilities during testing, but it also hinders the learning process by causing inefficient allocation of cognitive resources (Ashcraft, 2002; Cassady, 2004; Tse & Pu, 2012). When students experience anxiety, their cognitive resources are used for emotional regulation rather than for cognitive processing related to learning (Covington & Omelich, 1987; Hinze & Rapp, 2014). As a result, test-anxious students experience problems when trying to encode, organize and integrate new information, leading to incomplete mental models (Naveh-Benjamin et al., 1987).

In conclusion, test anxiety has negative effects on academic achievement because of anxiety blockage and negative effects on cognitive functioning. However, small levels of test anxiety might also positively influence learning by increasing concentration (Shernoff, Csikszentmihalyi, Schneider & Shernoff, 2003), motivation, and effort (Owens, Stevenson, Hadwin, & Norgate, 2012). Therefore, quizzing should be implemented in such a way that it does not induce too much test anxiety, for example by avoiding time limits (Aydin, 2010) and grading (Khanna, 2015). In the current study, test anxiety was measured immediately after the pre- and post- domain knowledge test. Though measuring test anxiety during the quiz (i.e. the learning process) would also be very insightful, it was decided not to do this in order to allow students to fully concentrate on the lecture content.

#### **2.4. Test anxiety in relation to quizzing**

Opinions on the effect of quizzing on test anxiety differ greatly. On the one hand, researchers argue that quizzing can be anxiety provoking and therefore hinder performance (e.g. Cassady, 2004; Nguyen & McDaniel, 2014). For instance, students might not feel ready to actively take a quiz about new material (Khanna, 2015), or experience an extra workload inducing anxiety (Chamberlain, Daly, & Spalding, 2011). Moreover, too complex quiz items might make students feel out of control, increasing test anxiety (Bledsoe & Baskin, 2014; Trevino & Webster, 1992).

On the other hand, research shows that frequent quizzing reduces test anxiety for 64% (McDaniel et al., 2011) or even 72% (Agarwal, D'Antonio, Roediger, McDermott, & McDaniel, 2014) of the students. Agarwal et al. (2014) hypothesize that test anxiety was reduced because students became familiar with taking tests. Another possible reason for reduction of test anxiety is found in a study by Wells and King (2006), which showed that metacognitive therapy leads to a significant decrease in worry, a dimension of test anxiety affecting academic performance (Cassady, 2004; Covington & Omelich, 1987). This suggests that the positive effect of quizzing on students' confidence levels, in turn, helps to reduce test anxiety. In line with this argumentation, Bledsoe and Baskin (2014) argue that quizzing reduces anxiety because it provides students with regular opportunities to check what they do and do not know, giving them a feeling of control.

Despite the difference in opinions, literature clearly shows that for quizzing to have no or positive effects on test anxiety, quizzes should not be graded (Hinze & Rapp, 2014), time limits should be avoided (Aydin, 2010), and multiple-choice questions are desirable (Zeidner, 1987). The quiz of the current study was designed accordingly, though some short-answer questions were included as well (see method). To investigate the effects of quizzing on learning outcomes, questions based on Bloom's taxonomy were used, as discussed in the following section.

## 2.5. Measuring learning outcomes

Learning outcomes in the current study were measured using a combination of low- and high-level questions based on the first four levels Blooms taxonomy. Low-level questions included *remembering* and *understanding* (i.e. knowledge in a similar situation), whereas high-level question focused on *applying* (i.e. knowledge in a new situation), and *analysing* (i.e. knowledge of elements and their relations) (Krathwohl, 2002). Based on previous research, it was expected that quizzing mainly improves test scores on low-level summative exam questions (Thomas et al., 2018; Agarwal et al., 2008). However, Carpenter (2012) suggests that quizzing can also promote performance on high-level questions.

Additionally, this study investigated the effects of quizzing on *delayed* learning outcomes, because students' understanding of educational content at the moment of quizzing differs from their understanding during the final exam due to decay, interference, or consolidation (Carpenter, 2012). For example, newly encoded information will be integrated with existing knowledge during students' sleep, which consolidates their understanding of the topic (Diekelmann & Born, 2010). In research however, little attention is paid to the effects of quizzing on longer retention intervals (McDaniel et al., 2011). In the current study, learning outcomes were measured two days after initial exposure to the educational content to include these long-term effects but still minimize the interference of external variables that may influence the study results.



### 3. Research questions and Hypotheses

Effective processing of recorded lectures is becoming essential due to the increasing pressure for secondary education to employ active learning strategies. Research suggest that quizzing can be used to improve delayed learning outcomes of these recorded lectures. Based on the presented theoretical framework, the following research question and hypotheses were formulated:

*What are the effects of quizzing in recorded lectures on pre-university students' test anxiety and delayed learning outcomes?*

**H1:** Quizzing in recorded lectures improves video engagement

Explanation: As mentioned by van der Meij and Dunkel (2020), students must engage with the recorded lecture effectively (e.g. watch the entire video, replay parts that are not understood) in order for the lecture to influence learning outcomes. In the current study, video engagement was measured using log files (see chapter 4).

**H2:** Quizzing in recorded lectures improves delayed learning outcomes

**H2.a:** Re-exposure to educational content leads to higher delayed learning outcomes.

**H2.b:** Active construction of knowledge leads to higher delayed learning outcomes.

**H2.c:** There is a correlation between level of confidence and delayed learning outcomes.

**H2.d:** The effects of quizzing are greater for low-level delayed summative exam questions compared to high-level questions.

**H2.e:** The effects of quizzing on delayed summative exam scores are greater for quizzed than for non-quizzed material.

**H3:** Quizzing in recorded lectures reduces students' test anxiety

**H3.a:** There is a correlation between level of confidence and test anxiety.

**H4:** Test-anxiety negatively influences the effects of quizzing on delayed learning outcomes

## 4. Method

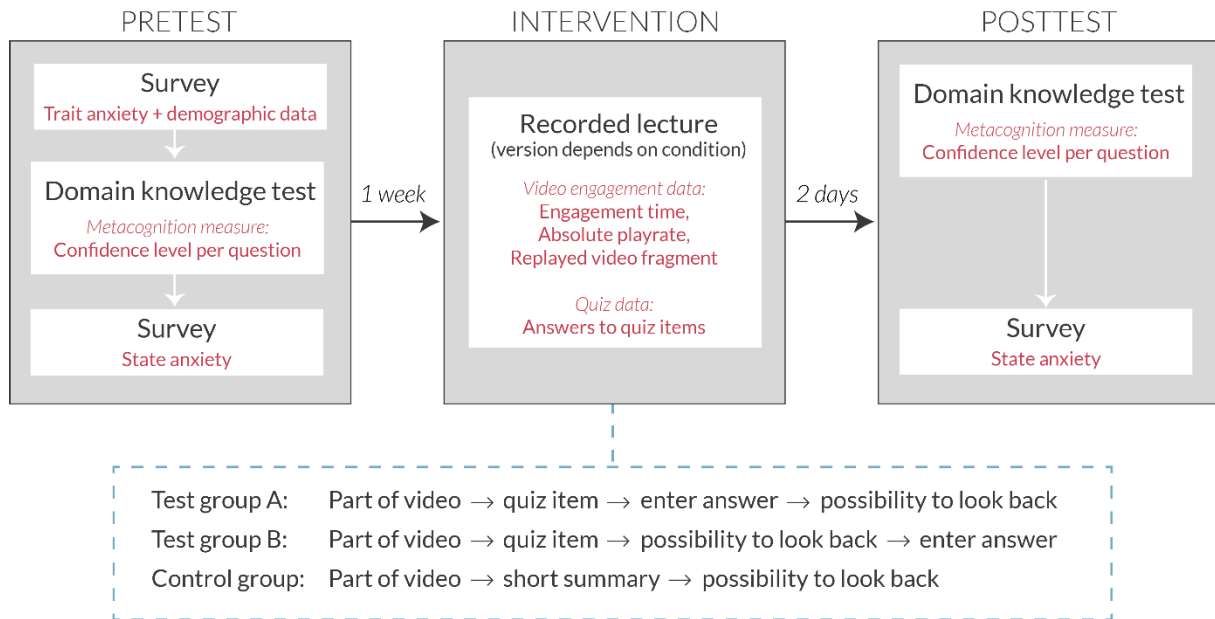
### 4.1. Participants & Design

A total of 70 pre-university students were included in this study. However, due to absence during one or more sessions, 21 dropped out. This means that the final sample included 49 pre-university students (65.3% female) from a Dutch high school that offers accelerated pre-university programs. Students ranged in age from 16 to 23 years ( $M = 18.80$  years,  $SD = 1.58$ ).

To answer the research questions, experimental research with a pre-posttest design was conducted. Data was collected using a test anxiety questionnaire, domain knowledge tests with self-reported confidence levels, and logfiles of a recorded lecture. The experiment contained two test groups (group A and B) and a control group (group C). Students were randomly assigned to one of the three test conditions. In the end, group A contained 16 students, group B included 15 students and 18 students were assigned to the control condition.

As indicated in Figure 1, students in all conditions received a segmented video, either interpolated by quiz items or by short summaries, and were allowed to re-watch a part of the video after the quiz or summary. The inclusion of short summaries in the control group is essential to investigate the extent to which the testing effect influences the learning outcomes compared to re-exposure effect. Moreover, by varying the structure of the test groups' videos, insights into the learning strategies prompted by quizzing as well as the effects on test anxiety could be obtained. Students in test group A were forced to actively construct their knowledge because they were not allowed to do a content check before answering the quiz items. It was expected that this improves learning outcomes but could also induce higher levels of test anxiety. Students in test group B were afforded a content check before answering the quiz item, which was expected to foster re-study and minimize test anxiety. It was expected that students from test group A would score higher in the post-test compared to test group B, because of the active construction of knowledge. In both test conditions, students received feedback once they submitted an answer and were allowed to re-watch the preceding video segment. It was hypothesized that students from both quizzing conditions would score higher on the domain knowledge test compared to the control group, due to more accurate confidence levels.

The experiment was divided over three moments of measurement (see Figure 1): 1) a pre-test to set a baseline for the confidence measure, test anxiety, and prior knowledge; 2) the intervention in which students watched the recorded lecture and answered the quiz items; 3) a post-test to analyse the effects of quizzing on measures of confidence, test anxiety, and delayed learning outcomes. Each appointment lasted for 50 minutes.



*Figure 1.* Overview of the research design. Grey boxes indicate moments of measurements; white boxes indicate instruments; the blue box specifies research conditions. The red text indicates when specified variables are measured.

#### 4.2. Instruments & Data analysis

**Pre- and post- domain knowledge test.** Offline domain knowledge tests were used to analyse the effects of quizzing on delayed learning outcomes. The tests contained both low-level as well as high-level questions (as discussed in Section 2.5). They were created in cooperation with a science teacher and were aligned with the theory discussed in the recorded lecture. Both tests were conducted at school, at similar timeframes. The pre-test contained twelve short-answer questions (Cronbach's  $\alpha = .67$ ), of which seven covered knowledge gained in previous chapters and five covered the to be learned material (see Appendix B). Example questions of the pre-test are “Write down the electron configuration of Calcium” (recap) and “Draw the Lewis structure of alcohol” (covered in the recorded lecture). The post-test contained twelve questions (Cronbach's  $\alpha = .71$ ) of which seven were also included in the quiz (see Appendix C). These seven questions covered the same content as questions in the quiz, but the level or context varied. For example, if a quiz question was “Which of the Lewis structures below is a correct representation of the carbonate ion?”, then the post-test question was “Draw the Lewis structure of the  $\text{CO}_3^{2-}$  ion”. For later data analysis, a distinction was made between low- and high-level questions (Cronbach's  $\alpha = .28$  and  $\alpha = .68$  respectively) and quizzed and non-quizzed questions (Cronbach's  $\alpha = .61$  and  $\alpha = .50$  respectively). In both tests, students could receive a total of 17 points. Feedback on the test results was not provided.

Alongside providing an answer to each question, students were asked to rate how confident they were of that answer on a scale of 0 to 1 (0 = not confident, 0.5 = semi confident, 1 = very confident). These confidence levels were used to analyse a dimension of students' metacognitive skills based on

two measures described by Huff and Nietfeld (2009): calibration accuracy (i.e. the absolute difference between expected and actual performance) and calibration bias (i.e. a measure of over- or underestimating performance). The actual performance was scored in the range of 0 (completely incorrect) to 1 (completely correct). Then, the calibration accuracy was calculated by dividing the sum of all absolute differences between expected and actual performance per question by the total number of questions. Calibration bias was calculated per question by subtracting the actual performance score from the reported confidence level. For example, if a student reported a confidence level of 0.5 for a correctly answered question, the calibration bias was  $0.5 - 1 = -0.5$ . This signed difference indicates that the student underestimated his/her performance for that question. The percentage of over-/underestimated questions was then calculated by dividing the number of over-/underestimated questions by the total number of questions in the pre- or post-test.

**Test anxiety questionnaire.** Students' level of trait and state test anxiety was measured using a paper questionnaire based on the STAI-A survey created by Bieling, Antony, and Swinson (1998). The STAI-A survey contains seven items measuring trait anxiety, which were extended by seven similar items measuring state anxiety (see Appendix A). Trait anxiety was only measured during the pre-test, whereas state anxiety was measured during the pre- and the post-test. The original questionnaire is based on a 4-point Likert scale. However, this was changed into a 5-point Likert scale to increase reliability and to allow participants to more accurately express their feelings (Lozano, García-Cueto & Muñiz, 2008). The questionnaire contained items such as "I worry too much over something that really doesn't matter" (trait) and "I felt nervous and restless when making the test" (state). For each item, participants rated to what extent they agreed with the statement, ranging from totally disagree (1) to totally agree (5). In addition, several background characteristics were collected during the pre-test, such as year of birth. Cronbach's'  $\alpha$  for the trait and state questionnaires was .83 and .88 respectively.

**Website.** A website was used to guide students through the recorded lecture and to collect relevant video engagement data (see section 'logfiles'). Students of the different research conditions visited different versions of the website, using their personal ID to login in. Figure 2 shows two screenshots of the website. The video segments and quiz items were alternately visible, guiding the user through the experiment as described in the research design. The user could play, pause, rewind and fast-forward the lecture as desired using the video controls. The quiz items appeared only at the end of the video, stimulating students to watch the video before continuing to the quiz. Moreover, students were obligated to answer each quiz item to stimulate active selection and organisation of information. The next section describes the content of the video segments and quiz items.

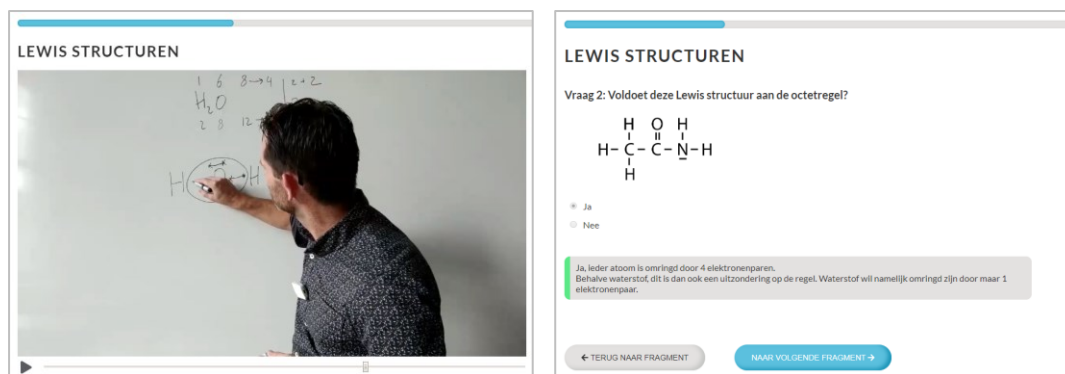


Figure 2. Screenshots of the intervention website. Left: video segment. Right: quiz item with feedback.

**Recorded lecture with quiz items.** To investigate the influence of quizzing as described in section 2.1, three versions of a segmented recorded lecture were created in cooperation with a science teacher. The control group received a simple video with interpolated short summaries (see Appendix D), in which rewinding was allowed. The two test groups received a video with quiz items (see Appendix E) and their opportunities to rewind were limited (see section 4.1). In each condition, students were allowed to work on the recorded lecture and quiz for a total of 50 minutes. The lecture was designed such that it could be finished well within this time limit to minimize the pressure being imposed on the students.

The recorded lecture contained five video segments of approximately four minutes each. First, the lecture's topic was introduced to prepare students for forthcoming information and thereby reduce test anxiety (Bledsoe & Baskin, 2014). Then, each video segment focussed on a specific topic, for example the process of constructing a Lewis structure. The segments were sequenced from simple to more complex and included both explanations of the topic as well as many examples.

After each video segment, students in the test conditions were provided with two or three questions about concepts discussed in the preceding segment(s). Students in the control group received short, textual summaries instead of quiz items. The quiz was created in cooperation with a science teacher to ensure that it contained questions commonly used in pre-university classes. The question format varied between multiple-choice and short-answer questions. Multiple-choice questions were used because they are perceived as less anxiety evoking (Zeidner, 1987). However, even though they are more anxiety evoking, short-answer questions were included in the quiz as well because these let students better reflect their knowledge (Anderson, 1987; Zeidner, 1987) which presumably yields more robust results compared to multiple-choice questions (Thomas et al., 2018). The quiz items were created based on the first four levels of Bloom's taxonomy. An example of a low-level question is: provide a definition for a given term. An example of a high-level question is: analyse an unfamiliar Lewis structure and define whether it is correctly drawn or not. Immediately after submitting their answer, students received feedback which could be used to correct errors and misconceptions (Fiorella & Mayer, 2015; Shapiro, 2009).

**Logfiles.** Logfiles (see Appendix F) were used to unobtrusively analyse students' video engagement and to collect their answers to the quiz items. Logs revealed, amongst other things, the time on task (i.e. engagement time) for each of the video segments. A high engagement time yields a greater learning effect because replays indicate that participants notice a need for better understanding and pauses most likely indicate reflection or study of the video contents (Meij & Dunkel, 2020). The engagement time was expressed as a percentage of the total video segment duration. For example, if a video segment contained 360 seconds and the student interacted with the video (interaction measures included replays and pauses) for 450 seconds, the engagement time was  $(450/360 \times 100 = 125\%)$ .

Another measure collected in the logfiles was the unique play rate, which indicated the percentage of the video that was watched by the student. This measure is important because students must watch the video for it to affect learning (Meij & Dunkel, 2020). Again, the unique play rate was expressed as a percentage of the total video segment duration. For example, if a student watched the first 50 seconds of a 200-second video, fast-forwarded to the end, answered the quiz and then replayed the last 60 seconds of the video, the unique play rate was  $(50+60)/200 \times 100 = 55\%$ . In other words, a unique play rate of 100% indicates that the student watched every second of the video segment at least once.

**Analysis of results.** User codes were employed to anonymously link data of individual students collected through multiple measurements. Incomplete datasets (i.e. if a student was absent during one or more session(s)) were excluded from the analysis. A check on random distribution of participants for gender, prior knowledge and trait anxiety revealed no significant difference between conditions.

Then, all variables were checked for normality using the Shapiro-Wilk test for small ( $n < 200$ ) samples. For normally distributed data, t-tests were used to analyse the differences between pre- and post-test scores. In addition, (repeated-measures) ANOVAs were used to analyse the effects of the different test conditions on video engagement, test anxiety and delayed learning outcomes. If data was not normally distributed, the non-parametric Kruskal-Wallis and Wilcoxon signed-rank tests were used instead. If these tests revealed a significant difference between conditions, ad hoc tests such as Dunn's comparison test were used to analyse the differences in more detail.

#### 4.3. Procedure

Teachers from a Dutch high school were asked to participate in the experiment with some of their classes. Once they agreed, they were asked for permission for approaching their students. In addition, the ethics committee of the University of Twente was asked for approval. Once permission was granted, a total of 70 pre-university students (partly) participated in the research project. Participants were sampled based on homogeneity to assure that large differences in students' level of knowledge would not affect the results of this study. Within the subset of pre-university students, convenience sampling was used to select participants who all attended the same classes, which was essential because

the intervention was part of a school's actual curriculum. Finally, the participating students were divided into three groups using purposeful random sampling. So, students in the same class did not participate in the same test condition *per se*.

Before the experiment, participating students were asked for consent after being informed about the purpose of this study. Also, to obtain valuable results, it was desirable that students were dedicated to this study. However, grading the domain knowledge test could reduce the expected testing effect, because students would probably study the quizzed and non-quizzed content equally well (McDaniel et al., 2011). Therefore, students who actively participated in the study were promised a bonus for their final course grade regardless of their scores on the quiz and domain knowledge tests. The measurements were conducted at the school location during school hours. See Table 1 for an overview of the procedure per specified session. Each session included approximately 60 students and lasted for 40 or 50 minutes.

Table 1

*Overview of the procedure per session*

	Pre-test	Intervention	Post-test
Duration	40 minutes	50 minutes	50 minutes
Instructions provided in advance	<p>Students receive a bonus for their final course grade regardless of their quiz/tests scores.</p> <p>Students are not expected to answer all the questions of the domain knowledge test correctly because content is partly new.</p>	<p>Tell students how the lecture will proceed, e.g. “the lecture consists of a few videos which are separated by quizzes you need to answer”.</p> <p>Tell students when they are (not) allowed to re-watch the previous video segment.</p>	<p>The bonus which students were promised does not depend on their score on the domain knowledge test.</p>
Procedure during session	<ol style="list-style-type: none"> <li>1. Students fill out informed consent</li> <li>2. Students fill out trait anxiety survey.</li> <li>3. Students complete the domain knowledge test.</li> <li>4. Students fill out state anxiety survey.</li> </ol>	<p>Using individual computers and headphones, students watch the recorded lecture and, if applicable, answer the quiz items.</p>	<ol style="list-style-type: none"> <li>1. Students complete the domain knowledge test.</li> <li>2. Students fill out state anxiety survey.</li> </ol>
Procedure after session	<p>Collect informed consent, surveys, and answers to the domain knowledge test and thank students for their participation.</p>	<p>Collect logfiles and thank students for their participation.</p>	<p>Collect surveys and answers to the domain knowledge test and thank students for their participation.</p>



## 5. Results

### 5.1. Distribution of demographics

Table 2 and 3 show the distribution of demographics among the three research conditions. ANOVA's showed that prior knowledge ( $F(2) = 1.69$ ;  $p = .196$ ) and trait anxiety ( $F(2) = 2.09$ ;  $p = .136$ ) were randomly distributed among the research conditions. A Chi-squared test showed that gender ( $\chi^2(2, N = 49) = 2.90$ ,  $p = .234$ ) was randomly distributed as well. However, age ( $F(2) = 5.17$ ;  $p = .009$ ) was not. A spearman correlation showed that age and post-test scores were highly correlated ( $r_s = .31$ ;  $p = .032$ ) but age and knowledge gain were not ( $r_s = .24$ ;  $p = .098$ ).

### 5.2. The effect of quizzing on video engagement

Table 4 shows that participants, on average, played 94% of each video segment (unique play rate) and engaged with the videos for 106% of the video length (engagement time). On average, videos were not watched more than once. The non-parametric Kruskal-Wallis test revealed no significant difference in engagement time ( $H(2) = 2.18$ ;  $p = .337$ ) and unique play rates ( $H(2) = 3.84$ ;  $p = .147$ ) between research conditions. The test did reveal a significant difference in the number of replays between research conditions ( $H(2) = 14.72$ ;  $p = .001$ ).

Table 2

*Distribution of demographics among the three research conditions*

Research condition	Age <i>M (SD)</i>	Prior knowledge (%) <i>M (SD)</i>	Trait anxiety <i>M (SD)</i>
Active construction (n=16)	18.44 (1.59)	39.80 (15.07)	2.60 (.92)
Re-exposure (n=15)	19.80 (1.66)	40.59 (16.55)	2.53 (.76)
Control (n=18)	18.28 (1.13)	31.13 (18.04)	2.06 (.82)
<i>Total (n=49)</i>	<i>18.80 (1.58)</i>	<i>36.85 (16.90)</i>	<i>2.38 (.85)</i>

*Note.* Active construction is research condition A, re-exposure is research condition B, control is research condition C.

Table 3

*Distribution of males and females among the three research conditions*

Research condition	Male (freq.)	Female (freq.)
Active construction (n=16)	8	8
Re-exposure (n=15)	5	10
Control (n=18)	4	14
<i>Total (n=49)</i>	<i>17</i>	<i>32</i>

Table 4

*Differences between video engagement measures in all three research conditions*

Research condition	Unique play rate (%) <i>M (SD)</i>	Engagement time (%) <i>M (SD)</i>	Replays (freq.) <i>M (SD)</i>
Active construction (n=16)	98.17 (2.92)	109.91 (11.59)	.03 (.10)
Re-exposure (n=15)	90.63 (25.54)	108.19 (25.41)	.35 (.55)
Control (n=18)	92.65 (24.33)	101.48 (33.92)	.00 (.00)
<i>Total (n=49)</i>	<i>93.84 (20.31)</i>	<i>106.29 (25.53)</i>	<i>.11 (.34)</i>

A Dunn-Bonferroni multiple comparison ad hoc test ( $\alpha = .017$ ) was used to further investigate the difference in replay behaviour between the three research conditions. This test showed a significant difference between the conditions of active construction and re-exposure ( $Z = 11.53$ ;  $p = .001$ ), and between the re-exposure and control condition ( $Z = -10.10$ ;  $p = .007$ ). No significant difference was found between the control and active construction condition ( $Z = 1.443$ ;  $p = 1.000$ ).

### 5.3. The effect of quizzing on learning outcomes

As can be seen in Table 5 and Figure 3, students in general scored higher on the post-test ( $M = 43.43$ ;  $SD = 20.16$ ) than they did on the pre-test ( $M = 36.85$ ;  $SD = 16.90$ ). A repeated-measures ANOVA was conducted to compare the test scores before and after the recorded lecture. There was a significant effect of time on test scores with  $F(1, 46) = 6.39$ ,  $p = .015$ ,  $d = .35$ . However, this effect did not significantly differ between the research conditions, with  $F(2, 46) = 1.40$ ,  $p = .256$ . Moreover, no significant interaction was found between time and research condition ( $F(2, 46) = 1.10$ ;  $p = .342$ ).

Table 5

*Differences in scores between all research conditions on the pre and post domain knowledge tests*

Research condition	Pre-test <i>M (SD)</i>	Post-test <i>M (SD)</i>
Active construction (n=16)	39.80 (15.07)	40.90 (16.04)
Re-exposure (n=15)	40.59 (16.55)	49.80 (21.45)
Control (n=18)	31.13 (18.04)	40.36 (22.17)
<i>Total (n=49)</i>	<i>36.85 (16.90)</i>	<i>43.43 (20.16)</i>

*Note.* Scores are presented as a percentage of the total amount of points students could obtain.

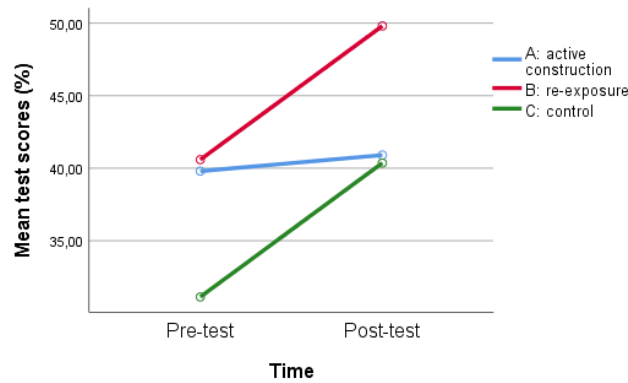


Figure 3. Mean test scores in the pre- and post-test for all research conditions.

### Performance on low and high-level post-test items

A Wilcoxon signed-rank test revealed no significant difference between the pre- and post-test low-level test scores ( $Z = 596.00$ ;  $p = .935$ ). As can be seen in Table 6, this means that students in general did not score differently on low-level pre-test questions ( $M = 38.20$ ;  $SD = 23.00$ ) than they did on low-level post-test questions ( $M = 37.48$ ;  $SD = 18.34$ ). The test did reveal a significant difference between the pre- and post-test high-level test scores ( $Z = 955.00$ ;  $p = .001$ ). This means that students in general scored significantly higher on high-level questions in the post-test ( $M = 48.12$ ;  $SD = 26.10$ ) compared to high-level questions in the pre-test ( $M = 35.66$ ;  $SD = 16.72$ ).

In addition, the *test score increase* was used as a measure to investigate the differences in the scores on high-level questions between the three research conditions. ANOVA test results were insignificant with  $F(2) = .567$ ,  $p = .571$ .

Table 6

*Differences between scores on low and high-level items in all three research conditions*

	<u>Pre-test</u>		<u>Post-test</u>		<u>Test score increase</u>	
	Low	High	Low	High	Low	High
Research condition	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Active construction (n=16)	41.99 (20.76)	37.85 (15.94)	35.21 (16.78)	45.39 (23.13)	-6.78 (22.63)	7.55 (22.64)
Re-exposure (n=15)	40.21 (24.69)	40.93 (14.30)	41.78 (18.42)	56.14 (27.32)	1.57 (20.87)	15.21 (18.83)
Control (n=18)	33.16 (23.83)	29.32 (18.03)	35.95 (19.95)	43.86 (27.44)	2.77 (25.93)	14.54 (25.16)
<i>Total (n=49)</i>	<i>38.20</i> <i>(23.00)</i>	<i>35.66</i> <i>(16.72)</i>	<i>37.48</i> <i>(18.34)</i>	<i>48.12</i> <i>(26.10)</i>	<i>-.72</i> <i>(23.32)</i>	<i>12.46</i> <i>(22.36)</i>

*Note.* Scores are presented as a percentage of the total amount of points students could obtain. Test score increase was calculated by subtracting the pre-test from the post-test scores.

Table 7

*Differences between scores on quizzed and non-quizzed items in all three research conditions*

Research condition	Quiz	Post-test	
	Total score <i>M (SD)</i>	Score quizzed items <i>M (SD)</i>	Score non-quizzed items <i>M (SD)</i>
Active construction (n=16)	66.88 (11.61)	55.15 (16.48)	26.65 (22.51)
Re-exposure (n=15)	62.99 (15.73)	63.92 (22.98)	35.69 (22.29)
Control (n=18)	n.a.	54.25 (30.83)	26.47 (21.57)
<i>Total (n=49)</i>	<i>64.94 (13.71)</i>	<i>57.50 (24.38)</i>	<i>29.35 (22.05)</i>

*Note.* Scores are presented as a percentage of the total amount of points students could obtain.

### Performance on quizzed and non-quizzed post-test items

Table 7 provides an overview of students' scores on the quiz and post-test items. The Wilcoxon signed-rank test revealed a significant difference between scores on quizzed and non-quizzed post-test items ( $Z = 24.50$ ,  $p < .001$ ). This means that students scored significantly better on quizzed questions ( $M = 57.50$ ;  $SD = 24.38$ ) compared to non-quizzed questions ( $M = 29.35$ ;  $SD = 22.05$ ). In addition, the non-parametric Kruskal-Wallis test was used to investigate the differences between the three research conditions based on the scores on quizzed and non-quizzed post-test items. The test revealed no significant difference between research conditions in either the scores on quizzed questions ( $H(2) = 1.26$ ;  $p = .532$ ) or the scores on non-quizzed questions ( $H(2) = 1.92$ ;  $p = .382$ ).

Finally, a Pearson correlation was used to investigate a possible correlation between quiz scores and learning outcomes. Neither the quiz score and the post-test score ( $r = .03$ ;  $p = .877$ ), nor the quiz score and the knowledge gain ( $r = -.23$ ;  $p = .250$ ) were strongly correlated.

### 5.4. The effect of quizzing on confidence levels

Table 8 and Table 9 provide an overview of students' confidence levels during the pre and post domain knowledge test. A repeated-measures ANOVA showed an insignificant effect of the recorded lecture (i.e. time) on calibration accuracy scores with  $F(1, 46) = .01$ ,  $p = .913$ . This means that students' calibration accuracy during the pre-test ( $M = .32$ ;  $SD = .10$ ) did not significantly differ from their accuracy during the post-test ( $M = .31$ ;  $SD = .10$ ). Moreover, this effect did not significantly differ between the research conditions, with  $F(2, 46) = 2.85$ ,  $p = .068$ . Also, no significant interaction was found between time and research condition ( $F(2, 46) = .27$ ;  $p = .766$ ). Finally, no correlation was found between calibration accuracy and knowledge gain ( $r_s = .18$ ;  $p = .209$ ).

In addition, students' calibration bias was analysed to gain more insight into the inaccurate calibration reported here. A Wilcoxon signed-rank test revealed no significant difference between the percentage of under- and overestimated test scores during the post-test ( $Z = 466.50$ ,  $p = .563$ ). This

Table 8

*Differences in calibration accuracy in all three research conditions*

Research condition	Pre-test accuracy	Post-test accuracy
	<i>M (SD)</i>	<i>M (SD)</i>
Active construction (n=16)	.28 (.10)	.29 (.09)
Re-exposure (n=15)	.35 (.08)	.33 (.10)
Control (n=18)	.32 (.12)	.32 (.11)
<i>Total (n=49)</i>	<i>.32 (.10)</i>	<i>.31 (.10)</i>

*Note.* Calibration accuracy is the absolute difference between students' predicted and actual test performance. A score of 0 represents high accuracy, a score of 1 represents low accuracy.

Table 9

*Differences in calibration bias in all three research conditions*

Research condition	<u>Pre-test</u>		<u>Post-test</u>	
	Underestimated <i>M (SD)</i>	Overestimated <i>M (SD)</i>	Underestimated <i>M (SD)</i>	Overestimated <i>M (SD)</i>
Active construction (n=16)	26.56 (14.34)	21.88 (11.33)	20.31 (13.25)	24.48 (17.86)
Re-exposure (n=15)	28.89 (16.63)	28.89 (14.73)	24.44 (14.59)	20.00 (11.27)
Control (n=18)	21.76 (17.18)	26.39 (21.44)	21.76 (14.89)	18.52 (18.86)
<i>Total (n=49)</i>	<i>25.51 (16.08)</i>	<i>25.68 (16.56)</i>	<i>22.11 (14.08)</i>	<i>20.92 (16.41)</i>

*Note.* The under- and overestimation scores are presented as percentages of the total number of questions.

means that students did not overestimate themselves ( $M = 20.92$ ;  $SD = 16.41$ ) more than they underestimated themselves ( $M = 22.11$ ;  $SD = 14.08$ ). Another Wilcoxon signed-rank test showed that the recorded lecture did not influence this for neither overestimation ( $Z = 154.00$ ,  $p = .065$ ) nor underestimation ( $Z = 475.00$ ,  $p = .231$ ) of test scores. Moreover, the non-parametric Kruskal-Wallis test revealed no significant difference between research conditions in either the underestimation ( $H(2) = .86$ ;  $p = .651$ ) or overestimation ( $H(2) = 1.69$ ;  $p = .430$ ) of post-test scores. Finally, no correlation was found between neither underestimation of test scores and knowledge gain ( $r_s = .26$ ;  $p = .069$ ), nor between overestimation of test scores and knowledge gain ( $r_s = -.17$ ;  $p = .233$ ).

### 5.5. The effect of quizzing on test anxiety

Table 10 provides an overview of students' anxiety levels during the pre- and post-test. A Wilcoxon signed-rank test revealed a non-significant difference with  $Z = 274.00$ ,  $p = .104$ . This shows that students, on average, did not feel more anxious during the pre-test ( $M = 1.70$ ;  $SD = .73$ ) than they did during the post-test ( $M = 1.57$ ;  $SD = .74$ ).

Table 10

*State anxiety during the pre- and post-test in all three research conditions*

Research condition	State anxiety pre-test <i>M (SD)</i>	State anxiety post-test <i>M (SD)</i>
Active construction (n=16)	1.76 (.76)	1.67 (.79)
Re-exposure (n=15)	1.85 (.63)	1.67 (.81)
Control (n=18)	1.52 (.79)	1.40 (.63)
<i>Total (n=49)</i>	<i>1.70 (.73)</i>	<i>1.57 (.74)</i>

*Note.* Scores are based on a 5-point Likert scale. A score of 1 represents no anxiety, a score of 5 represents high levels of anxiety

In addition, the non-parametric Kruskal-Wallis revealed no significant difference in test anxiety during the pre-test ( $H(2) = 3.77$ ;  $p = .152$ ) and during the post-test ( $H(2) = 1.73$ ;  $p = .422$ ) between research conditions. Finally, a Spearman correlation was used to investigate a possible correlation between confidence levels and test anxiety. Neither the post-test calibration accuracy and test anxiety ( $r_s = -.04$ ;  $p = .768$ ), nor the post-test calibration bias and the test anxiety ( $r_s = .02$ ;  $p = .881$ ) were strongly correlated.

### 5.6. The mediating effect of test-anxiety on learning outcomes

A Spearman correlation revealed a positive moderate, significant correlation between trait and state anxiety ( $r = .44$ ;  $p = .002$ ). Moreover, a negative moderate, significant correlation was found between knowledge gain and state anxiety during the post-test ( $r = -.36$ ;  $p = .012$ ).

## 6. Discussion

Using a pre- posttest design in a real but controlled setting (a lecture within an actual curriculum course) with different versions of a recorded lecture, it was found that quizzing did not influence pre-university students' test anxiety and delayed learning outcomes. However, this study did show that the recorded lecture in general was effective. The following sections describe the various researched effects of quizzing in more detail.

### 6.1. The effect of quizzing on video engagement

First, this study investigated the influence of quizzing in recorded lectures on video engagement, because students must engage with the recorded lecture effectively for it to influence learning outcomes (van der Meij & Dunkel, 2020). It was expected that quizzing improves video engagement (Hypothesis 1). Results showed that students, on average, watched 94% of the recorded lecture, which is promising. In addition, the average engagement time (i.e. playtime including pauses and replays) of 106% suggests that students paused and/or replayed parts of the video that were not understood. This data thus reflects effective video engagement. However, students who were provided with quiz items did not engage with the recorded lecture longer compared to students who were merely provided with short summaries. So, it can be concluded that short summaries are equally effective as quiz items in stimulating students to watch the entire lecture. However, this might change when students need to watch the recorded lecture at home because without the supervision of a teacher and/or researcher they might lack the motivation to watch it (Cardall, Krupat & Ulrich, 2008).

In contrast, students' replay behaviour did significantly differ among the research conditions. It was expected that students in the re-exposure condition would use the quiz items to recognise essential material (Nevid & Mahon, 2009) and use that to re-watch parts of the preceding video segment (Kovacs, 2016). Results confirmed this and showed that students in the re-exposure condition did significantly replay the same video segment more often compared to the other conditions. However, it should be noted that students in the re-exposure condition, on average, replayed a video only .35 times, which is surprisingly little. This suggests that most students did not use the quiz items to determine which parts of the video to (re-)watch but watched the entire video before continuing to the quiz.

Moreover, the overall lack of replays suggests that the textual feedback provided with each quiz item was sufficient for students to understand the material. This is in line with previous research, which showed that, especially after an incorrect response, providing feedback without the opportunity to take another look at the material significantly increased scores on delayed retention tests (Kang, McDermott & Roediger, 2007; Pashler, Cepeda, Wixted & Rohrer, 2005). However, the lack of replays also indicates that students did not take action to strengthen their comprehension of the lecture material (van der Meij & Dunkel, 2020), which infers that learning gains are not maximized. So, other measures besides quizzing should be taken to stimulate the desired replay behaviour. Kim et al. (2014)

recommended adding one-click access to essential parts in the video (e.g. by highlights on the timeline or by adding clickable links) such that students can easily go back and forth between important explanation steps. This possibly lowers the barrier for students to replay a video because they know where to look and time is not wasted on parts of the explanation they do not need. Another suggestion put forward by Shyr and Chen (2017) is to include scaffolds for students to plan and monitor their own learning. For example, they asked students questions at various points throughout the lecture, such as “What prior knowledge do you need to retrieve before watching the video?” or “Have you reached previously set goals?”. Their study showed that this approach significantly improved learning outcomes.

In conclusion, the majority of students watched (almost) the entire recorded lecture, which hints towards effective video engagement. However, though students who were allowed to look back at the previous video before answering the quiz also significantly replayed videos more often, the number of replays in all research conditions was virtually nil. Additional measures are needed to improve video engagement.

## 6.2. The effects of quizzing on learning outcomes

This study aimed to investigate the effects of quizzing in recorded lectures on students’ delayed learning outcomes. It was hypothesised that quizzing in recorded lectures improves learning outcomes (Hypothesis 2) because of re-exposure to content (Hypothesis 2.a) and/or active construction of knowledge (Hypothesis 2.b). Students, on average, did score significantly better on the post-test compared to the pre-test, suggesting that the lecture in general was effective. However, no significant differences in test scores were found between the quizzing and control conditions. So, the current study did not confirm the hypothesis that quizzing in recorded lectures improves delayed learning outcomes.

One of the reasons for this unexpected outcome might be the real classroom setting in which this research was set out. Even though this was an experiment in which factors such as extrinsic motivation and level and amount of practice were controlled, some factors might still have influenced the results. Examples of these factors are higher intrinsic motivation (McDaniel et al., 2011) and already implemented teaching methods (Nguyen & McDaniel, 2014). As discussed in the subsection *Performance on quizzed and non-quizzed material*, higher student motivation might have fostered effective learning strategies for both students in the control and the research conditions. In addition, the implemented teaching methods might have influenced the research results in two contrasting ways. On the one hand, if students already engaged in effective learning strategies (such as notetaking while watching the video), then quizzing might have been unnecessary (Nguyen & McDaniel, 2014). On the other hand, if students do not yet engage in effective learning strategies, forcing them to answer quiz items does not guarantee that they will put effort into using effective strategies rather than simple luck to do so (Bell, Simone & Whitfield, 2015).



Another possible explanation for the current results is that students in all research conditions had to complete a pre-test before watching the recorded lecture. This might have induced the reactive effect, which means that taking the pre-test lead to a change in students' behaviour (Yu, 2010). In the current study, this could mean that students used the questions in the pre-test to determine which parts of the video are essential, and thus, which parts they needed to pay attention to most. This priming effect might have reduced the effects of quizzing.

Moreover, it is noteworthy that even though the lecture had an effect score of .35 and improved students' test scores compared to the pre-test, students still scored relatively low on the post-test with an average score of 43%. An explanation for this striking result might be that, besides watching the recorded lecture once, students did not study for the low-stakes test. For example, in a study by Nguyen and McDaniel (2014) in which students were not allowed to study after completing the quiz, no testing effect was found either. Another likely cause is the selective attention induced by the summaries and quizzes in the recorded lecture. Previous studies already showed that learning effects of quiz items do not persist for untargeted information (Nguyen & McDaniel, 2014; Shapiro, 2009), which could explain why students only answered half of the questions correctly. The influence of selective attention on the effects of quizzing will be discussed elaborately in the subsection *Performance on quizzed and non-quizzed material*.

Altogether, the recorded lecture was effective, but quizzing per se did not improve delayed learning outcomes. Based on these results, it can be suggested that previously proven quizzing effects do not persist in a classroom setting once factors such as external motivation, the frequency and the level of practice are kept constant.

### **Performance on low- and high-level post-test items**

The learning outcomes were also analysed based on test scores on low- and high-level questions. It was expected that the effects of quizzing for low-level exam questions would be bigger compared to the effects on high-level exam questions (Hypothesis 2.d). The current study did not confirm this hypothesis. Results showed that students, in general, scored significantly better on high-level questions in the post-test compared to high-level questions in the pre-test. Similar results were not found for low-level questions. This suggests that the recorded lecture improved students' application and analysis skills, rather than their retention. In addition, no significant differences were found between research conditions.

One of the possible explanations for this striking result can be found in the lecture design. Throughout the lecture, the teacher focussed on application rather than factual information by repeatedly showing how to solve complex chemistry problems. Unfortunately, as the complexity of the material increases, the testing effect decreases (Gog & Sweller, 2015). So, for complex learning materials that are high in element interactivity, such as the worked examples of this study's recorded lecture, quizzing does not improve learning outcomes. However, the high-quality lecture did improve test scores for high-

level exam questions. This suggests that putting time and effort into designing a high-quality lecture might be equally or even more effective in terms of higher-order thinking compared to designing a high-quality quiz.

To summarize, the current study showed that quizzing per se does not improve scores on low-level exam questions. However, a high-quality lecture, either interpolated by quiz items or short summaries, can be used to improve scores on high-level exam questions.

### **Performance on quizzed and non-quizzed material**

To get a clearer picture of the effects of quizzing, Shapiro (2009) suggested to not only analyse the effects of quizzing on learning outcomes in general but also distinguish between scores on quizzed and non-quizzed material. It was expected that the effects of quizzing on summative exam scores are greater for quizzed than for non-quizzed information (Hypothesis 2.e). The current study confirmed this hypothesis and showed that students, on average scored significantly better on post-test questions about quizzed compared to non-quizzed material. This suggests that re-exposure and/or active construction of knowledge does improve summative exam scores. However, these effects do not persist for untargeted information, which is in line with previous studies (Mayer et al., 2009; Nguyen & McDaniel, 2014; Shapiro, 2009).

To begin with, some researchers link these results to students' video engagement during the recorded lecture. Nguyen and McDaniel (2014) suggested that quizzing influences students' behaviour *after* watching the video, as students only re-study parts of the lecture that were targeted by the quiz, neglecting other important information. Likewise, Kovacs (2016) showed that quizzing influences students' behaviour *while* watching the video, as students jumped to the quiz to see which material is essential and watched only those parts of the lecture in which this material is explained. Both strategies of selective attention could harm learning because students might miss out on keys ideas needed to create a coherent whole. However, the current study showed that students, in general, did not skip the video to see the quiz first, nor did they replay the video after seeing the quiz. This means that video engagement is not the reason why quizzing effects did not persist for untargeted information.

Another possible explanation might be found in the reasoning why quizzing effects occur in the first place. Research suggests that requiring students to make a choice about the correct answer to a quiz item, stimulates active organisation and integration of that information (Dunlosky et al., 2013; García-Rodicio, 2015). Later retrieval of this information during the summative exam then becomes easier. However, information that is not targeted by the quiz is not actively integrated and connections to that information are not reinforced. Retrieval of this information, therefore, might become more difficult, resulting in lower summative exam scores. Nguyen and McDaniel (2014) related these effects to a study by Anderson, Bjork, and Bjork, (1994), which even showed that retrieving *impairs* connections to related information.

Despite the compelling differences between scores on quizzed and non-quizzed material, no significant differences between research conditions were found. This suggests that re-exposure to content by the short summaries in the control condition was equally effective as re-exposure by the quiz. Shapiro (2009) suggested that quizzing is effective because quiz items prompt students to direct attention to essential material of the lecture. Summaries might have had similar effects (Leopold, Sumfleth & Leutner, 2013) which, in line with Shapiro's reasoning, could explain why no difference between research conditions could be found. Shapiro (2009), like many others (e.g. García-Rodicio, 2015; Nguyen & McDaniel, 2014), also mentioned the testing effect as another advantage of quizzing, as quiz items stimulate students to recall information short after encoding it. As mentioned before, intrinsic student motivation might have affected these effects. For example, when provided with a short summary, highly motivated students might have tried to recall related information from the video because of their motivation to understand the material, eliminating the need for extrinsic motivation induced by the quiz. This could be another reason why no differences in post-test scores between the three research conditions were found.

Altogether, the current study showed that students scored significantly better on quizzed or summarized material compared to non-quizzed or non-summarized material, possibly due to a lack of reinforcement, or even impairment, of connections to untargeted information.

### **The effect of quizzing on confidence levels**

Besides re-exposure to content and the testing effect, previous research listed metacognitive skills (i.e. knowing what you do and do not know about the material) as another reason for the effects of quizzing in recorded lectures. Amongst others, Dunlosky and Rawson (2012) stated that processing a recorded lecture might be problematic because students tend to overestimate themselves and therefore terminate learning before the information is learned well enough. Quizzing could be used to limit this behaviour. First, it was expected that students who were presented with quiz items throughout a recorded lecture would more accurately predict their performance on the summative exam questions compared to students who were given short summaries. However, the results showed no significant differences between those groups. Moreover, the calibration accuracy and calibration bias in the pre-test did not significantly differ from the calibration accuracy and bias during the post-test, suggesting that neither the summaries nor the quizzes helped students to more accurately predict their performance on the summative exam. A possible explanation for this might be the frequency of quizzing. The focus on *repeated* instruction in the study by Huff and Nietfeld (2009) suggests that in teaching students to identify where they lack knowledge using the quiz items and corresponding feedback, more frequent practice rather than a one-time application of quizzing is crucial.

In addition, it was expected that there was a correlation between confidence levels and learning outcomes (Hypothesis 2.c). For example, García-Rodicio (2015) stated that when students know where they lack knowledge, they can allocate additional cognitive resources to effectively process the provided

feedback and adjust their understanding of the topic. Similarly, Dunlosky and Rawson (2012) mentioned that students who underestimate their performance might benefit from overlearning. However, the current study revealed no such correlation. A possible explanation for this is that even though students might know where they lack knowledge (i.e. they do not overestimate themselves), they need effective self-regulatory skills to turn this into an actual increase in learning outcomes (Tan, Whipp, Gagné & Quaquebeke, 2018). Logfiles of the current study revealed that students barely replayed a video after receiving the feedback, which hints towards ineffective self-regulatory skills. Apart from this, the current study showed that students under- just as much as overestimated themselves, and the arising effects of under- and overlearning might have cancelled each other out.

In conclusion, the current study showed that quizzing in recorded lectures does not influence students' confidence levels, which are also not correlated to their learning outcomes.

### **6.3. The effect of quizzing on test anxiety**

Another goal of this study was to investigate the effects of quizzing on test anxiety. It was expected that quizzing in recorded lectures reduces students' test anxiety during summative testing (Hypothesis 3). First, it should be noted that in the current study, students' anxiety levels, in general, were very low. This can be explained by the fact that none of the tests were graded (Khanna, 2015) and students were provided ample time to finish the tests (Aydin, 2010). Then, research has widely proven the negative effect of test anxiety on students' learning outcomes (e.g. Ashcraft, 2002; Batchelor, 2015; Cassady, 2004)(Hypothesis 4). In line with previous research, the current study showed that test anxiety indeed negatively influenced students' knowledge gain. However, students in the current study, on average, did not feel more anxious during the pre-test than they did during the post-test. Also, no significant differences in test anxiety levels were found between the three research conditions. This suggests that quizzing neither reduces nor increases students' levels of test anxiety.

A possible explanation for this unexpected result is the frequency of testing. Agarwal et al. (2014) suggested that quizzing reduces test anxiety because students familiarize with taking tests. Students in the current study participated in one quiz only, which might not have been enough to elicit this described effect.

Moreover, studies reporting a reduction in test anxiety (Agarwal et al., 2014; McDaniel et al., 2011) used a general survey at the end of the academic year rather than a specific test anxiety questionnaire during or shortly after the summative exam. Test anxiety levels reported after such a long retention interval might be inaccurate, because of memory changes due to decay (Carpenter, 2012). Moreover, socially desirable answers caused by the explicit manner of questioning might have caused biased results (Arnold & Feldman, 1981). Based on more reliable research methods (anxiety levels were determined using multiple implicit questions answered immediately after the tests), the current study refutes reported effects of quizzing on test anxiety.

In addition, it was expected that there is a correlation between confidence levels and test anxiety (Hypothesis 3.a) because knowing what you do and do not know could lead to a decrease in worry (Bledsoe & Baskin, 2014; Wells & King, 2006). However, the current study did not reveal such correlation. This might be a logical consequence of the recorded lecture's lack of effect on students' confidence levels. Since students were not sure about what they did and did not know, they did not experience that feeling of control which was needed to reduce their test anxiety.

## **6.4. Implications**

### **Theoretical implications**

This study investigated the effects of quizzing on pre-university students' delayed learning outcomes. The majority of studies on the topic of quizzing were conducted in real classroom practices. Therefore, it was not clear whether learning outcomes increased because of quizzing or because of other variables such as the emphasis on to-be-learned material, higher student motivation, or the amount of practice. The current study was a more controlled experiment and showed that when factors such as external motivation and the amount of practice are kept constant, quizzing does not increase learning outcomes. This study thus adds to existing theories as it suggests that the widely reported positive effects of quizzing might have been influenced by factors other than the quizzing effect.

In addition, this study also investigated the effects of quizzing on test anxiety. Previous research has slightly touched upon the possibility that quizzing reduces test anxiety (McDaniel et al., 2011; Agarwal et al., 2014), but this has not yet been widely confirmed. In fact, Nguyen and McDaniel (2014) suggested that quizzing might even increase test anxiety. Also, studies on the topic of quizzing and test anxiety were conducted among various school levels, whereas the current study focussed on pre-university students specifically. So, this study adds to existing theories as it proved that quizzing neither increases nor decreases test anxiety among pre-university students.

### **Practical implications**

This study showed that a recorded lecture is effective in promoting learning and teachers can, therefore, use it as an effective teaching method. When doing so, the recorded lecture should be presented in small segments, each followed by a short recap (either a summary or a quiz) re-exposing students to the content discussed. Findings of the current study suggest that teachers should put time and effort into designing a high-quality lecture rather than designing a good quiz when the aim is to promote higher-order thinking. See the study of Brame (2016) for an elaborate discussion of effective recorded lecture design. However, if quizzing is presumed necessary, quiz items can be added without inducing test anxiety. Important to consider is that when quiz items are added to a recorded lecture, these should address the same content as the summative exam for the quiz to be effective.

### 6.5. Limitations

A potential shortcoming of this study might be the relatively short duration of the experiment. The recorded lecture only covered one lesson of a specific chemistry topic, rather than the entire topic. This short timespan and the lack of repeated lecturing and quizzing might have reduced the expected effects of quizzing. One of the alternatives was to prolong the experiment and include more lectures and quizzes to increase the expected effects. However, studies by Freilich (1989) and Urtel, Bahamonde, Mikesky, Udry, and Vessely (2006) which were spread over a period of 15 and 7.5 weeks respectively, reported similar results in which quizzing did not influence learning outcomes.

Another possible shortcoming of the current study is that students were asked to watch the recorded lecture in class rather than at home. The social presence of the researcher and teacher might have increased students' video engagement (Kop, 2011) and thus, possibly influenced the effects of quizzing. However, allowing students to watch the recorded lecture at home would mean that variables besides the ones of interest for this study would influence the results, which is highly undesirable.

### 6.6. Future research

First of all, this study showed that the widely reported quizzing effects in the real classroom do not persist for more controlled settings. Future research should investigate this in more detail and identify which factors of the real classroom setting do and do not influence the testing effect and in what way. For example, McDaniel et al. (2011) mentioned that grading possibly increases student motivation, which in turn influences the testing effect. Deeper insight into the factors influencing the quizzing effect will help to further improve the use of quizzing in educational contexts and facilitate the development of modern teaching methods such as the flipped classroom.

Moreover, this study showed that a high-quality lecture can improve higher-order thinking skills even though quizzing effects were nil, whereas other studies convincingly showed that quizzing can be used to increase learning outcomes (e.g. García-Rodicio, 2015; Mayer et al., 2009). Therefore, future research should investigate which of the two, a high-quality lecture or a high-quality quiz, is more effective in terms of increasing learning outcomes. These insights will help teachers to organize their time efficiently and remove possible uncertainties associated with implementing the flipped classroom concept.

In addition, the current study failed to reveal a correlation between quizzing and confidence levels and between quizzing and test anxiety. However, previous research does hint towards these correlations (e.g. Cassady, 2004; García-Rodicio, 2015; McDaniel et al., 2011; Nguyen & McDaniel, 2014; Szpunar et al., 2014). As suggested in the discussion of the results, extending the study to more than one recorded lecture and increasing the frequency of quizzing might add valuable insights regarding these correlations. Future research should, therefore, try to take away the ambiguities raised by the current study by replicating the study and extending it for a longer research period.

## **7. Conclusion**

In conclusion, in contrast to the expectations, quizzing in recorded lectures neither affects pre-university students' test anxiety nor their delayed learning outcomes. These findings suggest that the widely proven effects of quizzing do not persist for more controlled classroom settings. Moreover, the quality of the lecture seemed to overrule the effects of quizzing. Re-exposure to content in general (either by short summaries or quizzes) appeared to be effective, especially when targeting the same content as the summative exam.

The current study was one of the first to investigate not only the effects of quizzing in a more controlled classroom setting but also to incorporate the effects on test anxiety. Therefore, it adds high value to the existing research on quizzing and recorded lectures in general and results can be used to further improve the increasingly popular concept of flipped classrooms.

## References

- Agarwal, P. K., D'Antonio, L., Roediger, H. L., McDermott, K. B., & McDaniel, M. A. (2014). Classroom-based programs of retrieval practice reduce middle school and high school students' test anxiety. *Journal of Applied Research in Memory and Cognition*, 3(3), 131-139. Doi: 10.1016/j.jarmac.2014.07.002
- Agarwal, P. K., Karpicke, J. D., Kang, S. H., Roediger, H. L. & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology*, 22(7), 861-876. Doi:10.1002/acp.1391
- Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20(5), 1063–1087. Doi: 10.1037/0278-7393.20.5.1063
- Anderson, P.S. (1987). Comparison of student attitudes about seven formats of educational testing, with emphasis on the MDT multi-digit testing technique, presented at the Annual Meeting of the Mid-Western Educational Research Association, Chicago, IL, 1987. Chicago: ERIC.
- Arnold, H. J., & Feldman, D. C. (1981). Social Desirability Response Bias. *Self-Report Choice Situations. Academy of Management Journal*, 24(2), 377–385. Doi:10.5465/255848
- Ashcraft, M. H. (2002). Math anxiety: personal, educational, and cognitive consequences. *Current Directions in Psychological Science*. 11(5), 181–185. Doi: 10.1111/1467-8721.00196
- Attali, Y., & Powers, D. (2009). Immediate Feedback and Opportunity to Revise Answers to Open-Ended Questions. *Educational and Psychological Measurement*, 70(1), 22–35. Doi:10.1177/0013164409332231
- Aydin, S. (2010) The effects of young EFL learners' perceptions of tests on test anxiety. *International journal of Primary, Elementary and Early Years Education*, 40(2), 189-204. Doi: 10.1080/03004279.2010.510119
- Batchelor, J. (2015). Effects of clicker use on calculus students' mathematics anxiety. *PRIMUS*, 25(5), 453-472. Doi: 10.1080/10511970.2015.1027976
- Bell, M. C., Simone, P. M., & Whitfield, L. C. (2015). Failure of online quizzing to improve performance in introductory psychology courses. *Scholarship of Teaching and Learning in Psychology*, 1(2), 163–171. Doi:10.1037/stl0000020
- Bieling, P. J., Antony, M. M., & Swinson, R. P. (1998). The State-Trait Anxiety Inventory, Trait version: structure and content re-examined. *Behaviour Research and Therapy*, 36(7-8), 777–788. Doi:10.1016/s0005-7967(98)00023-0
- Bing, S. B. (1982). Role of adjunct questions and reading ability levels on rote and conceptual learning from prose. *Instructional Science*, 11(2), 129–138. doi:10.1007/bf00154883
- Bledsoe, T. S., & Baskin, J. J. (2014). Recognizing Student Fear: The Elephant in the Classroom. *College Teaching*, 62(1), 32-41. Doi: 10.1080/87567555.2013.831022
- Brame, C. J. (2016). Effective Educational Videos: Principles and Guidelines for Maximizing Student Learning from Video Content. *CBE - Life Sciences Education*, 15(4), es6. Doi:10.1187/cbe.16-03-0125
- Bruns, S. B. (2017). Meta-Regression Models and Observational Research. *Oxford Bulletin of Economics and Statistics*, 79(5), 637–653. Doi:10.1111/obes.12172
- Cardall S, Krupat E, & Ulrich M. (2008). Live lecture versus video-recorded lecture: Are students voting with their feet? *Academic Medicine*, 83(12), 1174–1178. Doi: 10.1097/ACM.0b013e31818c6902
- Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current directions in psychological science*, 21(5), 279-283. Doi: 10.1177/0963721412452728.
- Cassady, J. C. (2004). The impact of cognitive test anxiety on text comprehension and recall in the absence of external evaluative pressure. *Applied Cognitive Psychology*, 18(3), 311–325. Doi:10.1002/acp.968



- Chamberlain, S., Daly, A. L., & Spalding, V. (2011). The fear factor: students' experiences of test anxiety when taking A-level examinations. *Pastoral Care in Education*, 29(3). Doi: 10.1080/02643944.2011.599856
- Chi, M. T. (2009). Active-Constructive-Interactive: a conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1(1), 73-105. Doi: 10.1111/j.1756-8765.2008.01005.x
- Covington, M. V., & Omelich, C. L. (1987). "I knew it cold before the exam": A test of the anxiety-blockage hypothesis. *Journal of Educational Psychology*, 79(4), 393-40. Doi: 10.1037/0022-0663.79.4.393
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438-481. Doi: 10.3102/00346543058004438
- Dibattista, D. & Gosse, L. (2006) Test Anxiety and the Immediate Feedback Assessment Technique, *The Journal of Experimental Education*, 74(4), 311-328. Doi: 10.3200/JEXE.74.4.311-328
- Dickelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, 11(2), 114-126. Doi:10.1038/nrn2762
- 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(1), 4-58. Doi: 10.1177/1529100612453266
- Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: inaccurate self evaluations undermine student's learning and retention. *Learning and Instruction*, 22, 271-280. Doi: 10.1016/j.learninstruc.2011.08.003
- Fiorella, L., & Mayer, R. E. (2015). Eight ways to promote generative learning. *Educational Psychology Review*, 28(4), 717-741. Doi: 10.1007/s10648-015-9348-9
- Freilich, M. B. (1989). Frequent Quizzing, the Final Exam, and Learning: Is There a Correlation? *Journal of Chemical Education*, 66(3), 219-223.
- García-Rodicio, H. (2015). Questioning as an Instructional Strategy in Multimedia Environments: Does Having to Answer Make a Difference? *Journal of Educational Computing Research*, 52(3), 365-380. Doi: 10.1177/073563311557193
- Glass, A. L. (2009). The effect of distributed questioning with varied examples on exam performance on inference questions. *Educational Psychology*, 29(7), 831-848. Doi: 10.1080/01443410903310674
- Hinze, S. R., & Rapp, D. N. (2014). Retrieval (sometimes) enhances learning: performance pressure reduces the benefits of retrieval practice. *Applied Cognitive Psychology*, 28(4), 597-606. Doi:10. 1002/acp.3032
- Hong, E., & Karstensson, L. (2002). Antecedents of state test anxiety. *Contemporary Educational Psychology*, 27(2), 348-367. Doi: 10.1006/ceps.2001.1095
- Huff, J. D., & Nietfeld, J. L. (2009). Using strategy instruction and confidence judgments to improve metacognitive monitoring. *Metacognition and Learning*, 4(2), 161-176. Doi:10.1007/s11409-009-9042-8
- Kang, S. H. K., 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(4), 528-558. Doi: [10.1080/09541440601056620](https://doi.org/10.1080/09541440601056620)
- Khanna, M. M. (2015). Ungraded pop quizzes: test-enhanced learning without all the anxiety. *Teaching of Psychology*, 42(2), 174-178. Doi: 10.1177/009862831557314
- Kim, J., Guo, P. J., Seaton, D. T., Mitros, P., Gajos, K. Z., & Miller, R. C. (2014). Understanding in-video dropouts and interaction peaks in online lecture videos. In *Proceedings of the First ACM Conference on Learning @ Scale Conference* (pp. 31-40). New York: Association for Computing Machinery.

- Kop, R. (2011). The challenges to connectivist learning on open online networks: Learning experiences during a massive open online course. *The International Review of Research in Open and Distributed Learning*, 12(3), 19-38. Doi: 10.19173/irrodl.v12i3.882
- Kovacs, G. (2016). Effects of in-video quizzes on MOOC lecture viewing, *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*, 31-40. Doi: 10.1145/2876034.2876041
- Krathwohl, D. R. (2002). A revision of bloom's taxonomy: an overview. *Theory Into Practice*, 41(4), 212-218. Doi: 10.1207/s15430421tip4104\_2
- Leopold, C., Sumfleth, E., & Leutner, D. (2013). Learning with summaries: Effects of representation mode and type of learning activity on comprehension and transfer. *Learning and Instruction*, 27, 40-49. Doi:10.1016/j.learninstruc.2013.02.003
- Lohbeck, A., Nitkowski, D., & Petermann, F. (2016). A Control-Value Theory Approach: Relationships Between Academic Self-Concept, Interest, and Test Anxiety in Elementary School Children. *Child & Youth Care Forum*, 45(6), 887-904. Doi:10.1007/s10566-016-9362-1
- Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the Number of Response Categories on the Reliability and Validity of Rating Scales. *Methodology*, 4(2), 73-79. Doi:10.1027/1614-2241.4.2.73
- Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., . . . Zhang, H. (2009). Clickers in college classrooms: fostering learning with questioning methods in large lecture classes. *Contemporary educational Psychology*, 34(1), 51-57. Doi: 10.1016/j.cedpsych.2008.04.002
- 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(1), 40-49. Doi:10.1037/0022-0663.75.1.40
- McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B. and Roediger, H. L. (2013). Quizzing in Middle-School Science: Successful Transfer Performance on Classroom Exams, *Applied Cognitive Psychology*, 27(3), 360-372. Doi:10.1002/acp.2914
- McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L. (2011). Test-enhanced learning in a middle school science classroom: the effects of quiz frequency and placement. *Journal of Educational Psychology*, 103(2), 399-414. Doi: 10.1037/a0021782
- McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morissette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, 19(4-5), 494-513. Doi:10.1080/09541440701326154
- Naveh-Benjamin, M., McKeachie, W. J., & Lin, Y. (1987). Two types of test-anxious students: Support for an information processing model. *Journal of Educational Psychology*, 79(2), 131-136. Doi:10.1037/0022-0663.79.2.131
- Nevid, J. S., & Mahon, K. (2009). Mastery quizzing as a signalling device to cue attention to lecture material, *Teaching of psychology*, 36(1), 29-32. Doi: 10.1080/00986280802529152
- Nguyen, K., & McDaniel, M. A. (2014). Using quizzing to assist student learning in the classroom: the good, the bad, and the ugly. *Teaching of Psychology*, 42(1), 87-92. Doi: 10.1177/0098628314562685
- Nyroos, M., Schéle, I., & Wiklund-Hörnqvist, C. (2016). Implementing test enhanced learning: Swedish teacher students' perception of quizzing. *International Journal of Higher Education*, 5(4), 1-12. Doi: 10.5430/ijhe.v5n4p1
- O'Flaherty, J., & Phillips, C. (2015). The use of flipped classrooms in higher education: a scoping review. *Internet and Higher Education*, 25, 85-95. Doi: 10.1016/j.iheduc.2015.02.002.
- Owens, M., Stevenson, J., Hadwin, J. A., & Norgate, R. (2012). When does anxiety help or hinder cognitive test performance? The role of working memory capacity. *British Journal of Psychology*, 105(1), 92-101. Doi: 10.1111/bjop.12009

- 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(1), 3–8. Retrieved from <https://escholarship.org/uc/item/7fg8h6zq>
- Putwain, D. (2008). Do examinations stakes moderate the test anxiety–examination performance relationship? *Educational Psychology*, 28(2), 109–118. Doi:10.1080/01443410701452264
- Roelle, J., Roelle, D., & Berthold, K. (2018). Test-based learning: inconsistent effects between higher- and lower-level test questions. *The Journal of Experimental Education*, 87(2), 299–313. Doi: 10.1080/00220973.2018.1434756
- Shapiro, A. M., 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. *Computers & Education*, 111(C), 44–59. Doi:10.1016/j.compedu.2017.03.017
- Shapiro, A. (2009). An empirical study of personal response technology for improving attendance and learning in a large class. *Journal of the Scholarship of Teaching and Learning*, 9(1), 13–26.
- Shernoff, D. J., Csikszentmihalyi, M., Schneider, B., & Shernoff, E. S. (2003). Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly*, 18(2), 158–176. Doi: 10.1521/scpq.18.2.158.21860
- Shyr, W.-J., & Chen, C.-H. (2017). Designing a technology-enhanced flipped learning system to facilitate students’ self-regulation and performance. *Journal of Computer Assisted Learning*, 34(1), 53–62. Doi:10.1111/jcal.12213
- Suo, J., & Hou, X. (2017). A study on the motivational strategies in college english flipped classroom. *English Language Teaching*, 10(5), 62–67. Doi: 10.5539/elt.v10n5p62
- Szpunar, K. K., Jing, H. H., & Schacter, D. L. (2014). Overcoming overconfidence in learning from video-lectures: implications of interpolated testing for online education. *Journal of Applied Research in Memory and Cognition*, 3(3), 161–164. Doi: 10.1016/j.jarmac.2014.02.001
- Tan, F. D. H., Whipp, P. R., Gagné, M., & Van Quaquebeke, N. (2018). Students’ perception of teachers’ two-way feedback interactions that impact learning. *Social Psychology of Education*, 22(1), 169–187. Doi:10.1007/s11218-018-9473-7
- Thomas, R. C., Weywadt, C. R., Anderson, J. L., Martinez-Papponi, B., & McDaniel, M. A. (2018). Testing Encourages Transfer Between Factual and Application Questions in an Online Learning Environment. *Journal of Applied Research in Memory and Cognition*, 7(2), 252–260. Doi:10.1016/j.jarmac.2018.03.007
- Trevino, L. K., & Webster, J. (1992). Flow in Computer-Mediated Communication. *Communication Research*, 19(5), 539–573. doi:10.1177/009365092019005001
- Tse, C. S., & Pu, X. (2012). The effectiveness of test-enhanced learning depends on trait test anxiety and working-memory capacity. *Journal of Experimental Psychology: Applied*, 18(3), 253–264. Doi: 10.1037/a0029190
- Urtel, M.G., Bahamonde, R.E., Mikesky, A.E., Udry, E.M. & Vessely, J.S. (2006). On-Line Quizzing and Its Effect on Student Engagement and Academic Performance. *Journal of Scholarship of Teaching and Learning*, 6(2), 84–92.
- van der Meij, H., & Dunkel, P. (2020). Effects of a review video and practice in video-based statistics training. *Computers and Education*, 143. Doi: 10.1016/j.compedu.2019.103665
- Van Gog, T., & Sweller, J. (2015). Not New, but Nearly Forgotten: the Testing Effect Decreases or even Disappears as the Complexity of Learning Materials Increases. *Educational Psychology Review*, 27(2), 247–264. Doi:10.1007/s10648-015-9310-x

- Wells, A., & King, P. (2006). Metacognitive therapy for generalized anxiety disorder: an open trial. *Journal of Behavior Therapy and Experimental Psychiatry*, 37(3), 206-212. Doi: 10.1016/j.jbtep.2005.07.002
- Wigfield, A., & Eccles, J. S. (1989). Test anxiety in elementary and secondary school students. *Educational Psychologist*, 24(2), 159-183. Doi: 10.1207/s15326985ep2402\_3
- Wise, S. L., Plake, B. S., Eastman, L. A., Boettcher, L. L., & Lukin, M. E. (1986). The effects of item feedback and examine control on test performance and anxiety in a computer-administered test. *Computers in Human Behavior*, 2(1), 21-29. Doi: 10.1016/0747-5632(86)90019-1
- Yu, C. H. (2018). *Threats to validity of research design*. Retrieved February 13, 2020 from <http://www.creative-wisdom.com/teaching/WBI/threat.shtml>
- Zeidner, M. (1998). The Origins and Development of Test Anxiety. In M. Zeidner (Ed.), *Test Anxiety: The State of the Art (Perspectives on Individual Differences)* (pp. 145-170). Doi: 10.1007/0-306-47145-0\_6
- Zeidner, M. (1987). Essay versus multiple-choice type classroom exams: The student's perspective, *Journal of Educational Research*, 80(6), 352-258. Doi: 10.1080/00220671.1987.10885782

## Appendix A – Test anxiety survey

### A.1. Trait anxiety and demographics data

1	Wat is jouw gebruikersID?			
2	Wat is jouw geboortedatum?	Dag	Maand	Jaar
3	Wat is jouw geslacht?	Man	Vrouw	Anders
				Zeg ik liever niet
		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Beantwoord de onderstaande vragen door aan te geven in hoeverre je het eens bent met de stelling.

		Helemaal mee oneens	Beetje mee oneens	Neutraal	Beetje mee eens	Helemaal mee eens
1	Ik word gehinderd door een onbelangrijke gedachte	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2	Ik maak mij teveel zorgen over iets dat er niet toe doet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3	Ik word gespannen of onrustig wanneer ik aan mijn huidige zorgen en interesses denk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4	Ik heb verontrustende gedachten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5	Ik neem teleurstellingen zo serieus dat ik ze niet meer uit mijn gedachten kan zetten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6	Ik heb het gevoel dat moeilijkheden zich zo opstapelen dat ik ze niet meer te boven kan komen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7	Ik voel mij nerveus en onrustig	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## A.2. State anxiety survey

De volgende vragen gaan over de toets die je net hebt gemaakt. Beantwoord de vragen door aan te geven in hoeverre je het eens bent met de stelling.

		Helemaal mee oneens	Beetje mee oneens	Neutraal	Beetje mee eens	Helemaal mee eens
1	Tijdens het maken van de toets had ik het gevoel dat moeilijkheden zich opstapelden waardoor ik ze niet te boven kon komen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2	Tijdens de toets maakte ik mij teveel zorgen over iets dat er niet toe deed	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3	Ik voelde mij nerveus en onrustig tijdens het maken van de toets	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4	Tijdens het maken van de toets werd ik gehinderd door een onbelangrijke gedachte	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5	Tijdens de toets nam ik teleurstellingen zo serieus dat ik ze niet uit mijn gedachten kon zetten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6	Tijdens het maken van de toets werd ik gespannen of onrustig wanneer ik dacht aan mijn huidige zorgen en interesses	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7	Ik had verontrustende gedachten tijdens het maken van de toets	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Appendix B – Pre domain knowledge test with self-reported confidence levels**

Vul hier je gebruikersID in:

**Beantwoord** de volgende vragen in de daarvoor bestemde vakjes. Probeer alle vragen te beantwoorden. Het is niet erg als je een vraag fout hebt, sommige vragen gaan immers over stof die jullie nog niet hebben behandeld.

Geef ook bij elke vraag aan **hoe zeker je bent** van je antwoord.

Succes!

**Vraag 1** Wanneer spreken we van een moleculaire stof?

<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

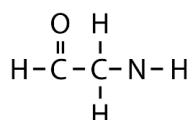
**Vraag 2** Hoeveel elektronen heeft een fosforatoom in de buitenste schil?

<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

**Vraag 3** Wat is de covalentie van zwavel?

<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

**Vraag 4** Kijk naar de onderstaande structuur formule. Klopt deze? Waarom wel/niet?



<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

**Vraag 5** Teken de structuur formule van HCN

<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

**Vraag 6** Geef de elektronen configuratie van Calcium

<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

**Vraag 7** Is er een structuurformule te tekenen voor SO<sub>2</sub> waarbij voldaan wordt aan de covalenties van de beide atomen? Zo ja, teken deze.

<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

**Vraag 8** Waarom staan de atomen in een H<sub>2</sub>O-molecuul onder een hoek?

<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

**Vraag 9** In de volgende structuur formule (koolstof monoxide) staan een aantal niet gedeelde elektronen niet getekend. Hoeveel losse elektronen bevinden zich nog bij elk atoom?



<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

**Vraag 10** Geef de elektronen configuratie van een aluminium-ion.

<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

**Vraag 11** Niet voor alle moleculen is een structuurformule te tekenen die in overeenstemming is met de covalenties van de atomen. Toch bestaan deze moleculen. Dat komt omdat er nog een andere voorwaarde is waar atomen aan wensen te voldoen. Welke voorwaarde is dit?

<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

**Vraag 12** Teken de Lewisstructuur van alcohol.

<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>



**Appendix C – Post domain knowledge test with self-reported confidence levels**

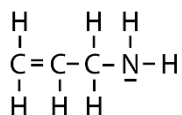
Vul hier je gebruikersID in:

**Beantwoord** de volgende vragen in de daarvoor bestemde vakjes. Probeer alle vragen te beantwoorden.Geef ook bij elke vraag aan **hoe zeker je bent** van je antwoord.

Succes!

**Vraag 1** Wat wordt bedoeld met de octetregel?

<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

**Vraag 2** Voldoet deze Lewis structuur aan de octetregel?☐

JA

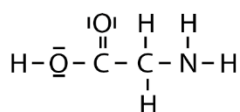
☐

NEE

Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>
---	-------------------------------------	---------------------------------------	-------------------------------------

**Vraag 3** Hoeveel elektronenparen bevinden zich in de buitenste schil van een molecuul  $\text{H}_3\text{PO}_4$ ?  
Dus: hoeveel streepjes mag je in totaal zetten?

<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

**Vraag 4** Klopt de onderstaande Lewis structuur? Zo niet, geef de correcte structuur.

<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

**Vraag 5** Teken de Lewis structuur van HCN

<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

**Vraag 6** Welke hoek hoort bij een 3 omringing?

<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

**Vraag 7** Hoeveel elektronen bevinden zich in de buitenste schil van het ion  $\text{SO}_4^{2-}$ ?

☐ 28      ☐ 30      ☐ 32      ☐ 40

Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>
---	-------------------------------------	---------------------------------------	-------------------------------------

**Vraag 8** Teken de Lewis structuur van het  $\text{CO}_3^{2-}$  ion.

<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

**Vraag 9** Juist of onjuist:  
Een molecuul heeft een formele lading wanneer een atoom in het molecuul niet voldoet aan de octetregel.

☐ JUIST      ☐ ONJUIST

Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>
---	-------------------------------------	---------------------------------------	-------------------------------------

**Vraag 10** Teken de formele ladingen op de juiste plaats in deze Lewis structuur van lachgas ( $\text{N}_2\text{O}$ ).

$\bar{\text{N}}=\text{N}=\bar{\text{O}}$			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

**Vraag 11** Waarom staan de atomen in een  $\text{H}_2\text{O}$ -molecuul onder een hoek?

<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

**Vraag 12** Laat zien dat er twee verschillende, maar gelijkwaardige, Lewis structuren te tekenen zijn voor het molecuul ozon ( $\text{O}_3$ ). Geef ook de formele ladingen aan.

<i>[provided space to answer question]</i>			
Hoe zeker ben je ervan dat je het juiste antwoord hebt gegeven?	Niet zeker <input type="radio"/>	Beetje zeker <input type="radio"/>	Heel zeker <input type="radio"/>

## Appendix D – Summaries used in the recorded lecture

### Video segment 1: introductie

Met structuur formules kan je niet altijd een molecuul op de juiste manier tekenen. Ook kan je de **vorm** van het molecuul hiermee niet bepalen. Dit kan je oplossen met een Lewis structuur.

Bij **Lewis structuren** kijk je niet alleen naar de gedeelde elektronen paren, maar ook naar **vrije elektronen** die zich in de buitenste schil van de atomen bevinden.

Voor het tekenen van de Lewis structuur gebruik je de **octetregel**: ieder atoom wil omringt zijn door 8 elektronen, dus 4 elektronen paren.

### Video segment 2: proces

Zo teken je een Lewis structuur:

1. Bepaal het totaal aantal elektronen in de buitenste schillen van alle atomen samen.
2. Bepaal het aantal elektronen dat nodig is voor om alle atomen aan de octetregel te laten voldoen.
3. Deel de elektronenaantallen door twee.
4. Bepaal het aantal gedeelde elektronenparen.
5. Teken de gedeelde elektronenparen.
6. Teken de vrije elektronenparen.

### Video segment 3: Voorbeelden

Je kan jouw getekende Lewis structuur controleren door het aantal **elektronen per atoom** te tellen en deze te vergelijken met het aantal elektronen dat dit atoom in de buitenste schil nodig heeft.

Er is sprake van een **ladingsverdeling** wanneer één van de atomen een elektron teveel heeft en een ander één elektron te weinig. Dit noemen we **formeel lading**.

### Video segment 4: ionen

Denk aan de **lading** van het ion bij het bepalen van het aantal elektronen paren.

- Als een ion een **positieve** lading heeft, is er één **elektron minder** aanwezig in de buitenste schil.
- Als een ion een **negatieve** lading heeft, is er één **elektron meer** aanwezig in de buitenste schil.

Door het aantal elektronen per atoom te controleren, kan je de **ladingen** van het ion op de juiste plek tekenen.

### Video segment 5: formele lading

Een **formeel lading** ontstaat wanneer een elektron van het ene naar een andere atoom overspringt om aan de octetregel te voldoen.

Je kan de Lewis structuur gebruiken om de **vorm** van een molecuul te bepalen:

- In een **drie omringing** staan de atomen onder een hoek van **120** graden.
- In een **vier omringing** staan de atomen onder een hoek van **109** graden.

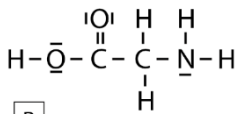
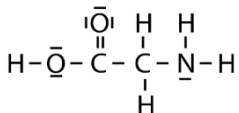
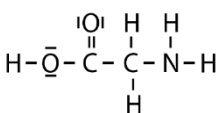
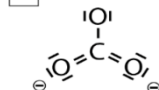
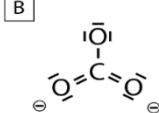
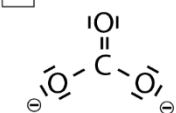
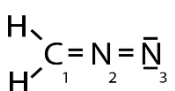
Soms kan je een molecuul/ion op **meerdere manieren** met een Lewis structuur weergeven. Het molecuul/ion is dan als het ware het gemiddelde van deze weergaves.

## Appendix E – Quiz items used in the recorded lecture

Table E1.

*Quiz items used in the recorded lecture.*

Nr	Type	Question	Feedback
1	MC	Welk van de volgende structuurformules is in overeenstemming met de covalenties van C, N, O en H? <div style="display: flex; justify-content: space-around; align-items: flex-start;"> <div style="text-align: center;"> <p><b>A</b></p> </div> <div style="text-align: center;"> <p><b>B</b></p> </div> </div>	Het juiste antwoord is B. Antwoord A is incorrect omdat het N-atoom rechts met 4 i.p.v. 3 gedeelde elektronenparen gebonden is. De covalentie van N is 3.
2	MC	Voldoet deze Lewis structuur aan de octetregel? <div style="text-align: center;"> </div>	Ja, ieder atoom is omringd door 4 elektronenparen. Behalve waterstof, dit is dan ook een uitzondering op de regel. Waterstof wil namelijk omringd zijn door maar 1 elektronenpaar.
3	MC	Via een aantal stappen kan je een Lewis structuur opstellen. Zet de onderstaande stappen op de juiste volgorde.  <ol style="list-style-type: none"> <li>1. Bepaal het aantal gedeelde elektronenparen.</li> <li>2. Teken de gedeelde elektronenparen</li> <li>3. Bepaal het aantal elektronen dat nodig is om alle atomen aan de octetregel te laten voldoen.</li> <li>4. Bepaal het totaal aantal elektronen in de buitenste schillen van alle atomen samen.</li> <li>5. Teken de vrije elektronenparen.</li> <li>6. Deel de elektronenaantallen door twee.</li> </ol> <div style="margin-top: 20px;"> <p>A. 3 – 4 – 6 – 1 – 5 – 2</p> <p>B. 4 – 3 – 6 – 1 – 2 – 5</p> <p>C. 4 – 3 – 1 – 6 – 5 – 2</p> </div>	B (4 – 3 – 6 – 1 – 2 – 5)
4	Open	Hoeveel elektronenparen bevinden zich in een molecuul H <sub>2</sub> SO <sub>4</sub> ? Dus: hoeveel streepjes mag je in totaal zetten?	16 elektronenparen (32 elektronen)

5	MC	Welke van de drie onderstaande Lewisstructuren is de enige correcte?	Het enige goede antwoord is A.  In structuur B bevinden zich 5 elektronenparen rond een van de O-atomen, en in structuur C mist een vrij elektronenpaar bij N.
		<div style="display: flex; justify-content: space-around; align-items: flex-start;"> <div style="text-align: center;"> <div style="border: 1px solid black; padding: 2px; margin-bottom: 5px;">A</div>  </div> <div style="text-align: center;"> <div style="border: 1px solid black; padding: 2px; margin-bottom: 5px;">B</div>  </div> <div style="text-align: center;"> <div style="border: 1px solid black; padding: 2px; margin-bottom: 5px;">C</div>  </div> </div>	
6	Open	Teken de Lewis structuur van Blauwzuur (HCN).	$\text{H}-\text{C}\equiv\text{N} $
7	Open	Welke hoek hoort bij een 4 omringing?	109
8	MC	Hoeveel elektronen bevinden zich in het ion $\text{NH}_4^+$ ?	Er bevinden zich 8 elektronen in $\text{NH}_4^+$ (houd rekening met de +: dus een elektron minder)
		a. 7 b. 8 c. 9 d. 12	
9	MC	Welke van de onderstaande Lewis structuren is een correcte weergave van het carbonaat-ion?	Alleen structuur C is correct.  In structuur A bevinden zich teveel of te weinig elektronenparen rond de O-atomen.  In structuur B staan de min-ladingen op de verkeerde plek (alleen het bovenste atoom heeft een min-lading).
		<div style="display: flex; justify-content: space-around; align-items: flex-start;"> <div style="text-align: center;"> <div style="border: 1px solid black; padding: 2px; margin-bottom: 5px;">A</div>  </div> <div style="text-align: center;"> <div style="border: 1px solid black; padding: 2px; margin-bottom: 5px;">B</div>  </div> <div style="text-align: center;"> <div style="border: 1px solid black; padding: 2px; margin-bottom: 5px;">C</div>  </div> </div>	
10	MC	Wanneer heeft een atoom een formele lading?	A
		a. Wanneer zich meer/minder elektronen in de buitenste schil bevinden dan gebruikelijk. b. Wanneer een atoom in een molecuul niet voldoet aan de octetregel. c. Wanneer de atomen in een molecuul niet in overeenstemming zijn met de covalenties.	
11	MC	In dit molecuul is sprake van formele ladingen. Welk atoom heeft hier een pluslading?	B
		 <div style="margin-left: 40px;">           a. Atoom 1            b. Atoom 2            c. Atoom 3         </div>	

**Appendix F – Logdata collected during the recorded lecture**

Table F1

*Variables collected during the recorded lecture using logfiles*

<u>Name</u>	<u>Description</u>	<u>Value</u>
userID	Personal ID used for all offline and online tests	2 letters + 3 numbers
userType	Condition the student participated in <i>A: video &gt; quiz &gt; answer &gt; possibility to re-watch</i> <i>B: video &gt; quiz &gt; possibility to re-watch &gt; answer</i> <i>C: video &gt; summary</i>	A / B / C
Item-[#]-pause	How often the video segment is paused in total	Frequency
Item-[#]-rewind	How often the video segment is rewound in total	Frequency
Item-[#]-forward	How often the video segment is fast-forwarded in total	Frequency
Item-[#]-engagement	Total time spend with video segment (including pause, rewind, fast-forward, and re-watch activities) as percentage of the total video segment length	Percentage
Item-[#]-playrate	Unique playrate; the percentage of the video that was actually watched by the student	Percentage
Item-[#]- skipVid	Whether the student jumped from the previous question to the next (Y), or watched (parts of) the video in between (N).	Y / N
Item-[#]-replay	How often the video is re-watched after seeing the quiz	Frequency
Item-[#]-replayedFrag	Which parts of the video were watched again after seeing the quiz	Segments in seconds
Item-[#]-answer	Answer to the quiz item as provided by the student	Answer
Item-[#]-correct	Whether the answer to the quiz question was correct (Y) or not (N).	Y / N

*Note.* The item numbers indicates the video segment (e.g. ‘item-2-pause’ represents the total amount of pauses in the second video segment).