University of Central Lancashire

Graeme Lamb

Using pre-written feedback in online tests to boost subsequent retest performance in adult students

MEd in Professional Practice in Education

EH4112 Dissertation
2017

Abstract

Twenty adult students in an evening class at a college of further education in the north-west of England took computer-based tests on a VLE (virtual learning environment). This took place for a consecutive period of six weeks on the subject of bookkeeping within a level 2 accounting course. Pre-written feedback generated for each question was given via the VLE, dependent on correct and incorrect answers to multiple-choice and short answer questions. Students were subsequently retested on the same material the following week in order to ascertain whether retest performance improved after the feedback intervention. 45% of students showed average increased performance on retests with other students showing no average change or decreased performance. 38% (30 out of 78) of individual retests showed an increase in performance from the original test. Students used resources of their own choice prior to retest, with some students using multiple revision resources and some students doing no subsequent revision. Student revision time prior to retest varied from no time to over an hour, with some correlation between methods used and time spent in feedback-led revision.

In some instances student performance decreased after revision, irrespective of time spent and number of methods used, suggesting ineffective revision techniques. Other potential factors include the working patterns and personal lives of participants. Students who scored higher on average on retest tended to work between six and eight hours per day, whilst students who worked eight or more hours per day were more likely to show no average change on retest or decreased

performance, suggesting that adult test performance in an evening class and the time to act upon feedback by the subsequent retest might be impacted by working patterns. It was also found that some students who worked in finance and accounting who had some pre-existing domain knowledge performed highly on both test and retest but also that the opposite was true in some students with pre-existing domain knowledge but who had no previous exposure to manual accounting methods.

It is suggested that performance of adults in evening classes in tests and exams and the factors relating to performance, particularly patterns of working and family life as well as pre-existing domain knowledge, are an under-researched area that would welcome further study to ascertain relevant factors that can increase or decrease performance.

Contents

Introduction

A key role of any tutor or teacher is the ability to enable students to close the gap between their current level of understanding and that which is required of them, in relation to the qualification or course being studied. Such gaps are often the difference between the domain knowledge with which a student begins a period of study (or their current level of understanding) and the learning outcomes set out by an awarding body that the student is expected to be able to demonstrate, through assessments, practical activities or exams, by the end of a period of study.

Research suggests that regular testing can improve retention and performance through what is known as the testing effect (Butler & Roediger 2007, Hattikudur & Postle 2011, Roediger et al. 2011, Carpenter 2012), and that testing when combined with formative assessment (sometimes referred to as Assessment for Learning) and feedback that is both appropriately given, received and acted upon (Sadler 1989 & 2013, Hattie & Timperley 2007, Hattie 2012) can enhance and deepen learning for students, closing the gap between what is known and what students are required to learn. The literature review within this study will consider the evidence of this claim in more detail, in particular considering the dialogic role of feedback (Nicol 2010 in Sadler 2013, Guasch et al. 2013) in the tutor and student relationship, the timing and content of feedback for best effect, the role of the student in appropriating feedback, the role of retrieval practice and testing in improving domain knowledge and performance, as well as some of the differing views on the role of feedback in improving the performance of students with greater or lesser

domain knowledge and how student perception of errors may affect the impact of feedback on learning.

Most of the research into feedback and testing has taken place in schools (Hargreaves et al. 2000, Clarke 2003, Roediger et al. 2011) or universities (Carrier & Pashler 1992, Roediger & Karpicke 2006, Butterfield & Metcalfe 2006, Carpenter & Kelly 2012) with the help of convenience samples of school students, undergraduates and postgraduates. This is entirely understandable given the rich availability of subjects in these sectors for researchers. However, an often-neglected area of research is that of adult students outside of a school or university setting. An example would be a college of further education, as such students often come to their learning at a different time of life with different priorities than (mostly younger) school or university students, and as such the internal and external factors that will affect learning, feedback and testing are likely to be different. Some have argued (Jarvis 2014) that although adult education was often seen as a field of practice within the university sector from the 1970s, that it was rarely seen as a field of study until recently, and as such there is space within the research literature on feedback and testing for more information with regards to adult education and the needs of adult learners.

The research study presented here is located within a college of further education and will examine many of the notions with regards to testing and feedback and when analysing findings and drawing conclusions will reflect on the specific nature of this sector. For consistency within this research study the terms tutor and student will be used rather than teacher and learner since it is the term commonly

used within the training provider in which this research took place, acknowledging the different relationship that exists between tutor and student (Rogers 2003) where adults may see the consequences and have expectations of their programme of study that younger students may not.

Research Aim & Objective

In some adult professional qualifications, such as the cohort of accounting students represented in this research study, there may be considerable variability of domain knowledge present at the beginning of the course. Among the twenty students who eventually completed this study, four of the students worked within accounting departments of larger businesses but without any accounting-related qualifications, and a further four students worked loosely in the areas of accounting and finance. The other twelve students either were not in employment or were not employed within accounting or finance but wished to in the future, and worked in a variety of careers such as fork-lift truck driver, blinds fitter and policeman. Such information is relevant in that it offers up assistance in terms of pre-existing domain knowledge that participants may bring in to the study that may influence test and retest scores, as well as recognising the diversity of a cohort of students of all ages. It has been argued (Boud 1989 in Usher, Bryant & Johnston 1997) that such learners bring their own valuable resources to learning with a greater knowledge of their own needs than (in particular) children.

Although larger (particularly undergraduate) classes exist, a class of twenty adult students with differing domain knowledge in a subject where each student brings something different to the learning process can provide a challenge for the tutor in terms of effective teaching, learning & assessment practice to enable all learners to progress and close the gap sufficiently to achieve the targeted learning outcomes throughout the course in a fast-paced qualification, where they undertake exams every seven weeks. The research study was deliberately positioned at a time in the course when students were studying a unit on bookkeeping, part of accounting with fixed rules and practices, and where questions could be asked which would have clear right or wrong answers.

A problem had been identified within the cohort that much knowledge was lost from week to week, even with simple recap exercises at the beginning of the next session. The question was asked by myself as the tutor and researcher as to what could be done about this situation, and a research question was devised and posed. Could pre-written feedback, in the form of feedback for both correct and incorrect answers for responses to questions in computer-based tests, provide a means of both bridging the gap between sessions and weeks in the classroom, whilst also being a method of improving performance and domain knowledge among this group of adult students? It was hoped that this feedback intervention would be able to catch and correct errors in students by the end of the week in which they had learned the material, as well as giving them the necessary guidance towards correcting the errors in a subsequent retest rather than just informing students whether they were correct or incorrect.

Literature Review

Introduction

Given that the research question in focus is the effects of pre-written feedback on student performance in subsequent retesting, it is appropriate that the primary focus of this literature review will be in the areas of feedback, testing and retrieval practice, including what the research literature has to say about the effectiveness and variability of such interventions. As well as feedback and testing, three further areas will also be referred to, although in less detail - memory, adult education, and some recent research and thinking on accounting. As indicated in the previous chapter, the subjects tested most frequently and on which most research has been based have often been school, undergraduate and postgraduate students, because of the rich availability of such research subjects. Whereas such research has provided invaluable data and theoretical understanding of the subjects, such research can sometimes lack data on the specifics of the participants under investigation, and the specifics of adult students in terms of their needs and what they bring to their studies are in comparison a relatively neglected area.

This literature review will conclude that although extensive research into testing and feedback has added a great deal of knowledge that is valuable for use in teaching and learning, that there are several caveats and particularities with regards to feedback and testing that mean it is not always easy to be certain what will work and what won't work in a given situation. For instance, there are issues with regards

to the relevance of feedback given, whether feedback is accepted by students, and what students subsequently do with any feedback given in terms of closing the gaps in their knowledge.

Feedback & Testing

Feedback in an educational setting can be seen in terms not only with respect to the content that is generated by a tutor but by the effect that it produces, whereby 'feedback is information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way' (Ramaprasad 1983, in Sadler 1989, p.120) where something can be defined as feedback if its intention is to provide information and guidance to the student in order to direct them in narrowing the gap. Others see feedback as that provided by an agent regarding performance of an action or task and therefore a consequence of performance (Hattie & Timperley 2007). Feedback is also seen, particularly within psychology for experimental researchers, in a more scientific form, where for example informative feedback refers to all post-response stimuli that were within the experimental design that comes to a student from outside with regards to performance (Narciss 2008), providing a more abstract approach to what might be understood by feedback.

It has been suggested (Kluger & DeNisi 1996) that much of the early research on feedback in the earlier parts of the twentieth century had rather suspect methodologies with very small sample sizes and feedback, and specifically feedback

interventions, were either increasing or decreasing performance due to a lack of a general theory at that time on the effects of feedback on performance. Many of these early studies had as few as four participants, reducing the significance of any findings in terms of generalisation. Sadler (1989) suggests that, with some exceptions, (Rowntree (1977), Bloom, Madaus and Hastings (1981), Black and Dockrell (1984) and Chater (1984) in Sadler (1989)) that the focus on high content validity in measurement and assessment in the past led to a deficiency of study and interest in the combination of feedback and formative assessment. High content validity here is understood as the close relationship between the means of measurement, for example the questions being asked, with regards to what is being assessed. Content validity may be particularly significant where learning providers are assessing students to external qualifications, where the needs and requirements of the awarding body must be met in terms of the relationship between forms of assessment and the learning outcomes of the qualification.

With a focus beyond content validity but also on the interplay between tutor and student in terms of feedback and formative assessment, the role of the tutor in this process must not be underestimated. Gipps saw the role of tutor assessment as being centred within a constructivist model of learning where knowledge and understanding is improved in learning by understanding what is known by the pupil and imbuing this knowledge with meaning (Gipps 1994). This leads to a shift of focus beyond the measurement of learning as an abstract concept, focused largely on the validity of the assessment environment as a means of measuring knowledge and progress, to an increased understanding of the role and needs of the learner and how they can progress with tutor assistance.

Black and Wiliam (1998) concur with this view, seeing the beginning of effective assessment of learners having as its starting point an understanding of what the learner does and does not know. For them, building on the work of Sadler, feedback has three elements - 'the *desired goal*, the evidence about their *present position*, and some understanding of a *way to close the gap* between the two' (Black & Wiliam 1989, p.10). This focused on moving beyond the giving of marks to pupils to feedback and assessment based upon meeting the learning needs of students, which might involve frequent short testing rather than occasional longer ones. Black and Wiliam further note the emphasis on constructive guidance on specific strengths and weaknesses being more significant than marks given to students (Black & Wiliam 1989, Black et al. 2002) where a learning opportunity may otherwise be lost due to the focus on marks. These works signify a shift around this in the thinking on feedback and assessment, and particularly a move towards a focus towards assessment for learning, (sometimes known as AfL) where assessment is a tool for both tutor and student on how the teaching and learning can be modified to improve learning, and not just a final summative assessment of what has been learned with no opportunity for redress or improvement. Ecclestone (2003) also recognises the significance of closing the gap and the role of learners in engaging with feedback, i.e. that feedback is not merely given but also received and acted upon pro-actively whereby 'without action or engagement, feedback remains a summative statement of achievement or of weaknesses' (Ecclestone 2003, p.53). Feedback, when used as part of learning, is formative in that it generates learning that enables further learning to take place (Stobart 2006).

Hattie (2012) suggests that feedback can, with an effect size of 0.79 listing it in tenth place out of one hundred and fifty influences on achievement, have a considerably larger effect than most other schooling methods, but that this comes with significant variability between different studies. Hattie sees four distinct levels on which feedback works - task, process, self-regulation and self, and the presence of the four feedback levels combined with the appropriate questions from the student, leads to the likelihood of greatly increasing the effectiveness of feedback, and feedback is likely to be better received by students if they feel that it is directly aimed at them rather than the whole class (Carless 2006 in Hattie 2012).

The inference is clear from authors already indicated and from others (Ecclestone et al. 2010, Sadler 2013), that the role of the student in processing feedback and then acting upon that feedback is of considerable importance in the likely success or failure of feedback interventions. Studies where feedback has been given but not acted upon (Anderson et al. 1971 in Metcalfe 2017) suggest that feedback has less impact than studies where students are given feedback based on errors that not only states the correct answer but why it is so, which is the case in the study presented here.

Some research has suggested that where high confidence errors are made, i.e. errors where the student had a high level of confidence that their answer was correct, the presence of post-test feedback that corrects the error and gives the correct answer has a higher chance of being successful in correcting the error (Butterfield and Metcalfe 2006). This is thought to be because the student giving a high confidence error has greater domain familiarity than a student giving a low

confidence error, who might in their own mind be guessing rather than calculating the answer, having insufficient knowledge to know the likelihood of their answer being incorrect or correct. The issue of the effect of subject familiarity (or domain knowledge) would seem to have an impact on how students view their performance, as others (Bereiter and Scardamalia 1993 in Sadler 2013) have argued that a certain level of subject familiarity in terms of a foundation of knowledge is needed before feedback can be useful and the student is able to process and act upon it. However, in contrast to this, other researchers suggest that feedback can be of more benefit to low achieving students than high achieving students, although this is dependent on clear instructions and clear direction for putting things right (Black & Wiliam 1989).

Some researchers and writers have noted that an important strand in whether feedback is likely to be effective or ineffective is the timing of feedback. Hattie (2012) suggests that feedback that is delivered too late may well mean the learning is stalled as students do not move on to more challenging tasks as they are awaiting feedback. Counter to this may be that students might have moved on to new material and the feedback for the old material, if not delivered in a timely and relevant manner, will be less useful as they will have moved on to new material and the opportunity for learning is lost or stalled. Worse still is that if corrective feedback is delayed, then students may continue making errors and those uncorrected errors may reinforce faulty learning or faulty domain knowledge.

Effective assessment and the role of feedback in assessment can be best described as a dialogue between the tutor and the students (Nicol 2010 in Sadler 2013) where assessment with feedback both informs the tutor on what is known and

what is unknown by the student whilst providing information for the student on closing the gap between where they are and should be, with effective feedback giving clear direction to the students and identifying errors and giving correct answers. It has been suggested (Kluger & DeNisi 1996 in Black et al. 2002) that feedback improved performance in approximately 20% of studies examined to that date, and that the danger was that feedback was seen or given as judgmental rather than developmental, in which cases it was less effective. One of the areas in which the giving and receiving of feedback may be productive for learners is after testing, and this is the context for the current research study.

The so-called 'testing effect' (Roediger & Karpicke 2006) is said to describe the benefit for learners of successful recall through testing of learned material. Many studies on testing have involved participants learning lists of words or images or the reading of texts which are then tested for recall, and are a common tool in both psychology and education. We have already considered the role of feedback in potentially improving learning and helping learners to bridge the gap, and this research study is an attempt to examine the role of feedback in the testing process. Significantly, some authors (Eisenkramer, Jaeger & Stein 2013) suggest that not only does testing have the benefit of providing information for the tutor and learner in terms of progress in learning, but that the process of testing itself through retrieval practice improves students' knowledge of the subject being studied, arguing that information retrieval is an attention-improving action. They suggest that the testing effect is made even more robust by the addition of feedback to the process (Butler & Roediger 2008, Smith & Kimbell 2010 in Eisenkramer, Jaeger & Stein 2013) and is more effective than simply re-reading course materials or student-created notes.

Others have argued that the format of the test (Kang, McDermott & Roediger 2007) is significant, arguing that short-answer quizzes provide more benefits than multiple-choice questions, in that research suggested that feedback after a short-answer quiz provided greater benefits in terms of subsequent performance than feedback after multiple-choice quizzes. They suggest that short-answer performance in testing was lower than in multiple-choice questions, and therefore the opportunity for the corrective feedback and its benefits was higher, especially when providing a correct response as feedback to an incorrect answer (Pashler 2005 in Kang, McDermott & Roediger 2007). There is also an argument (Roediger et al. 2011) that low-stakes quizzes are useful on a metacognitive level for learners in helping them to understand what they do and do not know. Research on 180 undergraduate students (Roediger & Karpicke 2008) suggested that testing, particularly after a period of days or a week, had more benefits on retention of materials than merely studying the materials. Some (Smith & Karpicke 2014) have suggested that multiple-choice questions can create problems by exposing students to errors and point towards research (Roediger & Marsh 2005 in Smith & Karpicke 2014) that suggests that this indeed is the case.

It is worth noting that issues of how students best retrieve information in a testing or examination environment are an important matter for awarding bodies to consider when setting exams for students in terms of how students can best show their knowledge and their learning and how assessment can be valid and reliable. It is also worth considering in terms of testing the potential negative effects within and around testing that can affect performance. A well-established and well-known

component of the testing process is the anxiety that some students commonly experience during a test or exam, particularly if there are high stakes. This may in some way lower the potential benefits of testing (Cassady & Johnson, 2002; Hembree, 1988; Lang & Lang, 2010; Ramirez & Beilock, 2011; Tobias, 1985 in Hinze & Rapp 2014) where disruptions in executive functioning or attentional control have limited and lowered performance, diverting learner resources away from the task in hand, with anxiety meaning that it is difficult for students to access their knowledge efficiently.

Memory

When considering issues such as test performance, the testing effect and retrieval practice, it is also important to briefly consider something of what the literature says about one of the key factors in human mental performance, the memory, because with retrieval practice the key component of the success or failure of it seems to lie in the effectiveness of otherwise of the memory. An exhaustive review of the literature on the subject of memory would no doubt cover too wide a scope here and occupy more space than is available. However, a brief overview of key concepts will it will be useful to focus on that are particularly relevant to testing and educational performance, and specifically to retrieval practice.

One psychological definition of memory is that it ‘is the process in which information is encoded, stored and retrieved. Via encoding, information from the world reaches our senses. Via storage, we maintain some of that information over

longer or shorter periods of time. And via retrieval we locate and return that stored information' (De Bruyckere et al. 2015, p.70).

Short-term memory, also sometimes known as working memory, is a vital component of performance in a test, particularly if that test takes place contemporaneously with learning that involves the encoding process that leads to storage and, for the purposes of our current study, retrieval. Distinctions have been made between learning and performance, but another suggested theoretical approach, advocated by Bjork & Bjork (Bjork & Bjork 1992, 2006) is that a distinction should be made between storage strength (the effectiveness of the encoding and storage) and retrieval strength (how effective a student is at then retrieving the knowledge being stored). This 'theory of disuse' is so named by Bjork & Bjork as they suggest that performance is a poor determinant of learning in such cases. In this model, information previously learned is not so much forgotten as unable to be retrieved. This is presented in contrast to Thorndike's 'law of disuse' (Thorndike 1914 in Bjork & Bjork 2006) where memory representations decay over time, and they point to other work (Hull 1943, Estes (stimulus-fluctuation theory) 1955 in Bjork & Bjork 2006) that suggest the importance of 'habit strength' in the ability to recall information and the important role of the control process of the rehearsal buffer (Atkinson & Shiffrin 1968) in information transferring successfully from short term memory and being recalled from the long term memory store at a later date. Working memory and the Atkinson-Shiffrin models of memory are the two predominant models in terms of the understanding in psychology of how human memory operates.

The impact of memory and aging has become more understood in recent decades, particularly with regards to longer-life spans meaning a greater prevalence of conditions such as dementia and Alzheimer's Disease. These conditions are something to be aware of in terms of adult education. It may be an issue for tutors in adult education that they have students who are studying courses for many reasons, but including the commonly-held belief that keeping the brain active can be helpful in reducing the risk of dementia.

Adult Learning & Accounting

It is recognised that the clear majority of undergraduate and postgraduate students are adults, and although many institutes of higher learning have a higher percentage of mature students than has been the case in past decades, there is a difference in a class of twenty-five adult students between the ages of eighteen and twenty-one than a similar class of students between the ages of twenty-two and sixty-five, which may be more typical in a college of further education.

Some have highlighted the wealth of experience from different areas that many adults bring into a learning environment, where some adult students will find themselves in a college of further education or similar institution having faced career-related rejection or marginalisation, which may come at the same time as a greater awareness of their failing body or memory (Roberson 2005). On the positive side, this rich resource of life experiences that adult learners bring to the learning environment can benefit themselves as well as others (Knowles 1980) and such

students are more likely to be looking to apply their learning directly to their current or intended career (Jarvis 1995) and have it individualized to their own context rather than pure, abstract learning.

Research into first-year accounting students (Byrne et al. 2014) suggested that students (particularly those who were male) were reluctant to seek help from each other, although the research did not state what effect this had on overall performance. Adult students in particular are more likely to have spent some time in employment, may bring in significant domain knowledge into the learning environment that enriches the learning (Jarvis 2014) and requires the tutor to find new and diverse ways of enabling the teaching and learning to help the diverse needs of different students. There may be a temporal difference in the relevance of feedback offered to younger students (who may only need this knowledge in the future) and older students already in the workplace in that sector, for whom the relevant domain knowledge required may be needed in the present.

A significant factor that has been recognised in terms of training future accountants has been the need for internalised, individualised learning that takes into account the changing nature of the profession (McDowall et al. 2015) particularly the increasing automation of lower level numerical tasks within the industry. Increasing computer capabilities and technological advances in computer storage capacity means that Artifical Intelligence (Economist 2017) is impacting on the industry in such a way, along with cloud computing, that much work that was previously undertaken manually by bookkeepers, such as entering transactions, can now be done automatically by computers. Software popular with small businesses,

such as FreeAgent and Xero is changing the landscape of what business owners can do without the help of bookkeepers and accountants, and is changing the future role of such professionals. The role of feedback in accounting courses in particular has been recognised in terms of equipping accounting students for the next generation of work available to them to the extent that 'assessment involving interesting and challenging assessed learning tasks that are individualised, authentic and with regular formative and summative feedback, in the context of an integrated set of interventions (Teamwork, Teacher-Student Relationship and Instruction), can support a large proportion of students to experience deep learning in an accounting course' (Turner & Baskerville 2013, p.594).

Conclusions

The role of feedback and testing have been extensively researched, both in laboratory and educational environments, and there is more than sufficient evidence to show that both can have positive results. However conflicting research results in some cases shows that this is not universally true and the presence of certain conditions makes feedback and testing more likely to be successful along with the benefits of retrieval practice and the testing effect.

Feedback is seen to have its most positive results when it is task-based and focused on what the student must do in order to improve their performance and close the gaps in their knowledge. It is also important for feedback to be given only a brief time after the relevant exercise for students to be able to act upon it in a timely

and relevant manner before moving on to other tasks. There is some debate over whether feedback is at its most effective for low-achieving students and students with varying levels of domain knowledge. The role of the student, who is free to accept, reject or modify feedback, is a crucial factor in the overall process, which is best seen as a process of dialogue between tutor and student.

The testing effect is well-documented and there is a clear benefit to retrieval practice, which research suggests is of greater benefit with short-answer questions than with multiple-choice questions as multiple-choice questions may expose students to errors. Testing shows clearer benefits to performance over re-reading course materials or studying them again.

Much of the research into feedback and testing has taken place in institutions of higher education such as universities, as opposed to the current study which is focused in a college of further education, and adult students who are older bring a variety of experiences and domain knowledge to the learning environment that younger students may not and which may impact on overall performance. Feedback as part of assessing accounting students as part of individualised learning may help such students to deepen their learning, at a time when the industry itself is undergoing significant change due to increasing automation of lower level bookkeeping tasks.

Methodology

Introduction

In this chapter, the methodology of the research will outlined and discussed. Firstly, I will be examining the research strategy in relation to the key research questions and exploring the methodological imperatives underpinning this research with a consideration of the various theoretical assumptions, information about participants and sampling and an exploration of the ethical issues raised by and during this research. The latter part of this chapter will then outline the research methods used, including issues of validity and reliability and also a short consideration of some of the limitations of the research methods used, including the implications of these methods for findings, conclusions and recommendations.

Research Strategy & Research Question

The key question to be asked during this research was with regards to the potential effect of pre-written feedback on subsequent test performance in an online testing environment. Could pre-written feedback, offered to students immediately after the completion of a test, guide students towards achieving a higher score in a subsequent retest as an indication of improved learning. This would involve students taking a short computer-based test lasting ten to fifteen minutes, after which they would be provided with pre-written feedback for both correct and incorrect answers, taking a retest the next week on the same topic. The computer-based tests were to be taken on the VLE (virtual learning environment) of the training provider, where

questions, answers and feedback would be stored. Feedback was individually written for each question by myself, and would comprise of different feedback for both correct and incorrect answers. Representative examples of some of the questions asked and feedback provided can be found in Appendix C.

The design of the study was that of a broadly mixed-methods approach. This involved the predominant methodological paradigm (in terms of data gathering) being a quantitative, positivist stance, with most data within the study being of a numerical nature, and an interpretivist element later in the study to attempt to explain the test/retest data gathered in a constructivist search for the meaning behind the results. This mixed-methods approach allowed for the greatest possibility of accurate data that could present a wider picture (Denscombe 2008 in Cohen et al. 2011) of what was taking place within the study, beyond the mere test scores.

The first part of the study involved six test/retest cycles providing numerical data of test and retest scores, and then relatively short questionnaires designed to gather information from participants that would hopefully help to provide information relating to exploring and explaining the test/retest data. It was considered important to have both quantitative data for statistical analysis and qualitative data for trying to explain the deeper narrative behind the test results, specifically in terms of telling the story of this cohort of students in response to the testing and retesting process. Qualitative data can help in this way to explore the human or personal side of the story behind the raw statistical data, which on its own may not answer the key research questions that have been asked without assistance (Churches & Dommett (2016: 5-7). Specifically, in the context of this study it was considered important to

ask participants about what they did with the feedback that they were given. This was because in trying to ascertain the significance of feedback in improving test scores, an understanding of how each student used feedback was considered vital. Further questions were added during the study as potential lines of inquiry grew that asked participants about their working patterns, noting that was this was a group of adult students attending their training provider in an evening class, after many had spent a day at work and had come to class straight from their workplace.

Theoretical Assumptions

The theoretical assumptions underpinning this research are in the areas of testing, including retrieval practice, and the use of feedback in improving learning and performance. It has already been discussed within the literature review that several studies suggest that retrieval practice in the form of short-answer and multiple-choice quizzes can be said to be more beneficial than mere studying (e.g. re-reading of notes) of a topic again. The assumption behind the research question was that feedback, given in an appropriate and timely manner, would help the students in correcting errors and in retrieval practice performance in subsequent retests. The basic structure within the first part of this research study (during the first six weeks) was:

1. New material studied (Tuesday & Wednesday, in class)

2a. Test (Wednesday, in class)

2b. Feedback given along with results (Wednesday, in class)

3. Revision (out of class, done by student)

4. Retest (in class the following Tuesday, prior to study of new material)

This process, running from Wednesday to the following Tuesday, comprised one test/retest cycle. It is not universally agreed in the research literature that feedback is always beneficial to students. Rather, as previously discussed, the methods, quality and timing of the feedback are significant, as well as how that feedback directs students to ways in which they can improve their performance and deepen their learning on a specific subject. The feedback used in this study was designed to be a combination of elaborated (in order to regulate learning and emphasise key points) and corrective/verification feedback (Guasch et al. 2013) that lets the student know the correct answer. Feedback was written in such a way that participants, whether they answered correctly or incorrectly, would be informed of not only the correct answer but also why it was correct.

Participants & Sampling

The cohort chosen for study was that of twenty-three students studying for their first (level 2) vocational qualification in accounting, after which they would have a basic grounding in bookkeeping and other key accounting concepts. This qualification was offered by the AAT (Association of Accounting Technicians) and is an acknowledged first step into accounting and bookkeeping. This group of students, eventually reduced to twenty students because of health factors or unemployment changes, was a group of adult students between the ages of twenty-two and fifty-

seven, attending the learning provider (a college of further education) for two three-hour lessons twice a week, on Tuesday and Wednesday evenings at 6.00pm.

This group is a type of non-probability sample known as an opportunity or convenience sample. This type of sample is often used in educational research, as previously mentioned, due to the relative ease of access to students in terms of gathering data. However, a drawback of using this type of group, that may not be selected for any specific criteria other than availability, is that the researcher must be careful to acknowledge that 'it does not represent the wider population; it simply represents itself' (Cohen et al. 2011, p.155). Therefore, when analysing subsequent data and drawing conclusions the researcher needs to be aware of the non-generalizable nature of the information across a wider population and in not drawing any firm conclusions without recognizing the limitations (Bell 2005). There were no incentives offered to potential participants to encourage them to take part in the study, but it was explained that it was hoped that this study would be of use to them in their learning as well as for future student cohorts, in gaining a deeper understanding of feedback and testing with adult students.

Ethical Considerations

There were a number of ethical issues to be considered within this research prior to commencing study, including informed consent from participants, confidentiality & data protection, and the relationship between the researcher, who was also the course tutor, and the research subjects.

Prior to the commencement of the research, written consent was sought and received from the learning provider. Once this was given, verbal consent was sought from the students with a brief overview of how the research was to take place and what would be involved in the process in terms of time, commitment and expectations. An ethics proposal was submitted (see Appendix A) and once this was accepted, formal written consent for participation (see Appendix B) was gathered from students. All students consented at that point and were reminded that they were able to withdraw from the study at any point should they so wish. As already mentioned, three participants left the course for various reasons, and their consent was presumed to be withdrawn at that point. Informed consent was used with a consideration of the guidelines on voluntary informed consent produced by the British Educational Research Association (BERA 2011) where participant students understood and agreed to their participation as well as their right to withdraw. The fact that written consent was not gathered from the participants until after the proposal was accepted meant that there was a period of four weeks between the students having the project explained to them and giving initial verbal consent and their final, written consent. This meant that students could think about their participation and were not coerced into participation in too short a period of time, allowing informed and thoughtful consideration by participants to take place prior to consent.

Given the nature of the opportunity sample, in that it was a class of students taught solely by the researcher who was first and foremost the course tutor, it is important to consider the nature of the relationship between tutor/researcher and

students, more specifically the power dynamic of the relationship. It is not uncommon in teaching for there to be a power dynamic between tutor and student where the tutor is in a natural position of power over the student. This manifests itself at its most basic level as the student being in the classroom environment as defined and set out by the tutor or learning provider and following the tutor's lesson plans and schemes of learning. In short, the tutor is in charge. This can be increased if the tutor is involved in providing summative assessment marks, for example grading a final assignment, then it may increase the power in the hands of the tutor. Age can also be a dynamic, in particular when the tutor is an adult and the students are children or teenagers. In the case of this research group, all the students were adults, which can to some extant minimised this power dynamic, and yet the tutor is still in charge of the lessons and the direction of the teaching, learning and assessment strategies. However, an important question in this study in terms of power and responsibility involved was based around the right to withdraw and students feeling coerced into participating either to please or to avoid displeasing the tutor. No students, excluding those who left the course, withdrew from the study, although I do not feel that students would have felt uncomfortable about withdrawing should an issue arose, and there is no evidence that students felt coerced into taking part in the study.

It was also important to consider issues of data protection and confidentiality. Responses to test/retest questions were initially stored within the VLE where the testing took place, and at this stage answers were viewable against a participant's name. Due to the nature of the computer software (Moodle) used, it was not possible to fully restrict access to this data, and tutors who had access to course modules, which was four tutors including the researcher, would be able to see that data.

Crucially however, participants would only be able to view their own responses and results rather than those of other participants, and only when they were logged into the system on their own personal login given by the training provider. Any data that was taken off the training provider's internal computer network, for example spreadsheets containing results, scores and analysis, was anonymized prior to being stored elsewhere to maintain confidentiality of personal information, whereby each participant was assigned a numerical code for cross-referencing. Examples of these unique IDs can be seen in the next chapter when data analysis and findings are discussed. When data was stored elsewhere, it was stored on a password-protected memory stick or a password-protected software cloud provider (Microsoft OneDrive) with no personal data identifiable. Once data had been analysed and stored safely, the original responses were removed from the training provider's virtual learning environment. There was no need within this research study to ask significant personal questions of participants within the study, although in later questionnaire participants were asked basic questions about their working patterns.

Research Methods

When designing the research study, it was initially considered that rather than taking the original test in classroom conditions, that students would take the test at home on their own computers. This approach was rejected for several reasons. The first reason was the possibility that not all students would have suitable computers at home on which to take the test, and this turned out to be the case. Action research during teacher training had highlighted in a small group of mentoring students that

only eight out of ten students had personal computers at home, with an increasing number of students using tablet computers (Lamb 2016). A second reason for rejecting the idea of students doing the test at home was that it would then be more difficult to control external factors during the test, and that the attention to the test required by students might be much harder in their home environment for personal and family reasons, such as not being able to give full attention to the test because of family demands. It was also possible that due to many of them having busy lives that they would not do the test at all. It was important to note prior to beginning the study whether individual students would need assistance in accessing the VLE in order not to be disadvantaged in the test, and all participants were shown how to access and then demonstrated accessing the VLE as well as taking a pilot practice test. It has been noted that the 'adult characteristics of self-directedness and computer self-efficacy' (Knowles 2014, 169) are important when designing and using computer based instruction and testing for adults. It was not presumed that participants would have the required skills prior to the research study, but there was no evidence that students needed extra help. After observing the students in other computer-based exercises, it was not considered necessary to give further instruction on general computer usage, but before the first test, the nature of how the tests would be completed and the practical steps were demonstrated to the whole class.

The advantages of students taking the test in the classroom included the ability for myself as researcher to see their responses straightaway, to know which students had finished the test and which were still answering questions, and to control the environment in terms of heat and light to keep the test environment

conditions as similar as possible. Students in the class sat in the same seats every week, in a room that was set out as a computer room, so that they took every test and retest on the same computer. As students had already sat one computer-based exam for their course by this point, and several smaller quizzes (not part of this research study) on the same computers and in the same environment on the VLE, they were not unfamiliar with the process of computer-based testing by the time that the research study commenced.

Test questions were written and pre-written feedback was written at the same time as the questions. Questions were based upon the learning expected for that session, and were written in style as closely as possible, allowing for the limitations of the question-writing module in the VLE, to those expected of them in their subsequent external exam. For example, if they were likely to be asked to calculate payroll costs and assign them to different accounts, inputting their answer in a box or using drop-down multiple-choice boxes, they would be asked the question in a similar format with a similar method of response expected. In designing the questions for this study, I was aware of previous work (Pashler 2005 in Kang, McDermott & Roediger 2007) that suggested particular benefits for corrective feedback in short-answer questions as opposed to multiple-choice questions, and so where practical short-answer options were used for asking questions although for some questions it was more appropriate to use multiple-choice questions or matching type questions, as these were likely to be similar to question styles that participants would encounter in a genuine exam situation.

After the test feedback was available to them immediately on screen until the end of the evening class. Test questions for the retest would test the same knowledge, but not using the same questions. For example, using the above example of payroll costs, in the retest students would be expected again to calculate payroll costs and assign them to the correct accounts, but would be given different numbers and calculations in order to achieve this. A short-answer question testing knowledge in the original test would also be tested with a short-answer question in the retest, and a multiple-choice or matching question in the original test would also be tested in the retest with a similar type of question.

After the completion of the six test/retest cycles, further questions were asked on the working patterns of the students, remembering that this was an evening class of adult students, many of whom were coming to lessons after a full day at work. Students were asked about their feedback storage methods, i.e. how they noted the feedback that they were given after the original test and prior to the subsequent retest from several available options. They were also asked what methods they used to revise between the test and retest as well as how much time on average they spent revising each week for the subsequent retests from a range of available options. It was hoped that the answers to questions on working patterns and the use of feedback would give valuable qualitative information to help explain and interpret in some way the results from the test/retest cycles. The questions from these questionnaires, originally completed online, are replicated in appendix D.

Validity & Reliability

The reliability of this research study is reliant partly on, first and foremost, whether the experiment could be replicated with relative ease by other researchers. In terms of the test and retest scenario involving the questions and the feedback, this is quite simple. Indeed, as the experiment is set up on a learning provider VLE, it would be easy for another tutor to run the experiment with another group in the same or a different classroom with almost no extra effort, as the questions and feedback could be simply exported from one VLE using the Moodle software and imported into another. It would be possible, although with some effort, to convert the tests and retests from an online to an offline (pen and paper) format, although some consideration would have to be given to how feedback was mediated to students in this scenario. This method could also be applied relatively easily using different questions and answers for different academic subjects where appropriate.

Within the subject of bookkeeping, and the design of the questions asked of participants, the answers are clear for each question in that they are objectively true without human judgment or subjective opinion being part of the answer, therefore each answer is clearly either correct or incorrect. In terms of validity, ensuring that all students took the tests and retests at the same time, in as consistent environmental settings and situations as possible (for example, heating and lighting) help the internal validity. Each student was only able to access the tests by logging onto the system with their own personal login, and the results for each student were logged on the system against their personal login, and therefore it was possible to ensure that the results for each student were recorded accurately for that student.

Environmental and internal variables were controlled as much as possible, certainly much more so than the original research design plan of having students complete tests and retests in their home environment.

External validity, to the extent that generalisable findings can be extrapolated in to other situations is less certain, due to the size of the sample used, and the within-design nature of the study. For this study there was no control group and all students were subject to the same internal variables in the study, although there will undoubtedly have been some external variables at work due to the nature of the research as well as the participants. These external variables would have included: working patterns of participants, tiredness or illness of participants on the day, health conditions (one student was in the third trimester of pregnancy during the study) and also the amount or effectiveness of work done after the first test ready for the subsequent retest. These are external factors over which the researcher will have little or no control, but must be considered in the context of the results of the research and any findings and conclusions that can be drawn, and will be discussed at some length in the next chapter.

Limitations of Methods Used

Recognised limitations in this research study include the type and size of the research sample. As an opportunity sample of only twenty students and with sample sizes for test/retest cycles being as low as eight, there must be exercised inevitable caution with regards to the conclusions that can be extrapolated from the results. One recent study on student self-assessment versus individualised instructor

feedback had forty-nine participants (Gibbs & Taylor 2016) whereas other research on testing and memory has had sixty (Carrier & Pashler 1992), eighty (Smith & Karpicke 2014) or even one hundred and eighty students (Roediger & Karpicke 2006).

The fact that this was an opportunity or convenience sample also adds limitations in that we cannot presume that the results gathered are generalisable to a wider population. This does not mean that any results gathered are meaningless or unimportant, and findings and conclusions may raise further questions for subsequent research as well as highlight areas of general interest within the further education sector, particularly regarding teaching adults. It is also possible that the length of the study was a limitation in terms of gathering a large quantity of data. A longer length to the study, perhaps ten to twelve weeks rather than six weeks might have provided information that was more stable than that gathered in a relatively brief period. As mentioned above, there is no control group within this research, which might have been useful in terms of exposing students to different variables as part of the research.

A final limitation was the inability within the software environment in which the testing took place for students not to receive marks, as marks were an integral part of the quiz module that could not be hidden. Therefore, students would receive a mark for their test as well as feedback. Black & Wiliam (1989) suggest that guiding students about specific strengths and weaknesses is preferable without marks being given. It was therefore possible within this research study that students might focus more on the marks received rather than the feedback received, even though students were

regularly asked to focus on the feedback rather than the marks and they were told at the beginning of the study that the significant part of the study was the use of feedback.

Data Analysis & Findings

Introduction

The major source of data for this study is quantitative, referring to the test and retest scores of participants, consisting of six test/retest cycles over a six-week period. Considering the central research question of whether pre-written feedback would improve scores in subsequent testing, inevitably these scores are an essential part of the data gathering process and the first to be analysed. A key point to note is that data was only analysed for tests and retests when participants were present for a full test/retest cycle, consisting of a test one week on a Wednesday and a retest the following week on a Tuesday. If a participant was present for the test but absent for the retest, that data has been excluded from the final analysis.

However, it was recognised early in the research process that qualitative data would be essential, both for triangulation and to begin to analyse the results gathered. As a result of this, after the completion of the initial phase of research, participants were asked several questions which it was hoped would shed light upon the data collected and enable more rigorous findings and conclusions.

Data Collection & Analysis

Twenty-three participants began the study, three left the course during the research period without completing a test/retest cycle, leaving twenty students.

Although all participants were asked to attend all sessions whenever possible, participant sizes varied between eight and sixteen for each test/retest cycle, with a major factor for this variance being illness and family issues. It is worth remembering that many of these students were also employees, employers, wives, husbands, partners and parents (in one case a parent-to-be), combining their studies with busy lives at home and work. Altogether between them the participants took part in seventy-eight test/retest cycles, making a total of one hundred and fifty-six tests and retests within the study (table 1).

Table 1: Overall test/retest averages for all 6 test/retest cycles, including percentage change on retest & number of participants for each cycle

	Test	Retest	% difference	Participants
Test 1	80.00	77.90	*-2.10*	14
Test 2	81.30	80.00	*-1.30*	8
Test 3	85.70	92.80	+7.10	14
Test 4	83.80	86.30	+3.00	16
Test 5	78.20	80.50	+2.30	11
Test 6	86.00	83.30	*-2.70*	15

Test/Retest Scores

In terms of whole class average performance, for three of the test/retest cycles performance was higher on retest than test and for the other three test/retest cycles performance on retest was lower than on the original test. Mean average performance for initial tests varied by 7.8% between the lowest (test 5) and highest (test 6). Mean average performance for retests varied by 14.9% between the lowest (retest 1) and highest (retest 3). The closest difference between test and retest

performance (test/retest cycle 2) of -1.30% was also the group with the smallest sample size of only 8 participants.

Out of the seventy-eight retests that took place across all test/retest cycles with all participants, thirty (38.46%) showed an increase in performance from the original test, twenty-four (30.77%) showed no overall change in performance on retest and a further twenty-four (30.77%) showed a decrease in performance on retest. Therefore, although more retests showed an increase in performance than either no overall change or decreased performance, the quantity of retests that showed an increase in performance was under 40% of the total and more retests in total showed no overall change or a decline in performance, suggesting minimal benefits.

The mean average overall test score was 82.50% and the mean average retest score was 83.47%, indicating a small but not significant increase of <1% in performance between tests and retests when averaged across all participants. It was notable that the first two test/retest cycles had both shown an average decrease in performance on retest with 50% (eleven out of twenty-two) of the participants in the first two cycles showing a decreased score on retest. It is not possible to fully explain this, although a factor might have been a lower level of familiarity with the assessment environment, even though participants had taken part in similar quizzes within the VLE prior to the commencement of the study. The next three test/retest cycles all showed an increase in performance on the retest. Twenty out of forty-one retests across these three cycles showed an increase in performance, and the retest in the third cycle had the highest number of participants (nine out of fourteen or 64.29%) increasing their performance on retest, which will most likely explain the

overall increase on retest for that cycle of +7.10%. Perhaps surprisingly, the final test/retest cycle showed once again an average decrease in performance on retest, with six out of fifteen (40%) of participants showing decreased performance on retest. However, it is worth nothing that the final test/retest cycle also had the highest mean average score (86%) in the initial test, which may have been a factor.

Table 2: Average participant scores by percentage sorted by students who increased, no overall change, and decreased on average

	Students who increased on average on rest	**Students with no overall average change on retest**	**Students who decreased on average on retest**
Average Test	81.56%	83.00%	85.33%
Average Retest	89.22%	83.00%	78.54%

Participants who increased their scores on retest were likely to perform poorer on the original test than those whose scores were poorer on retest. The nine participants who on average scored higher on the retest scored 81.56% on initial testing whilst the six participants who scored lower on the retest scored on average 85.33% on the original test. The group of five participants who showed no overall change in their retest scores found their test/retest scores to be somewhat in between the average scores of the other two groups. These results are similar to findings by others, (Butler & Roediger 2007) who suggested in their research that the lack of impact of feedback for some students was due to the higher performance on the initial test.

It is worth noting that out of the five participants who showed no overall change in their test/retest scores, two of the participants had identical test/retest

scores across all cycles whereas the other three participants showed some variance in that sometimes their retest score was higher or lower than the original test, but their average score across all cycles was the same for both tests and retests.

Once the data for individual test and retest performance is analysed, a more nuanced picture begins to emerge. This data can be seen in table 3 (below).

Table 3: average test/retest scores and average performance organised by test/retest cycle participation

Test/Retest cycles	No. of participants	Total participants	Average test	Average retest	% Incr/Decr	Increase (overall)	No overall change	Decrease (overall)
6	3	3	82.89	82.44	-0.45	1	0	2
5	5	8	81.40	89.80	+8.40	5	0	0
4	5	13	84.50	82.25	-2.25	1	2	2
3	5	18	84.67	81.33	-3.34	1	2	2
2	2	20	80.00	87.50	+7.50	1	1	0

It can be seen from table 3 that participation among students was varied. Three participants took part in all six test/retest cycles, with five students taking part in five cycles, five students taking part in four cycles, five students taking part in three cycles and two students taking part in only two cycles. Therefore, the majority (thirteen out of twenty) of students took part in at least four test/retest cycles. It is unclear as to why participants who took part in in five test/retest cycles show a more marked tendency to increase or decrease on retest, with all five participants increasing on average on retest. It may be related to attendance or effort in revision, but that is harder to argue when considering the fact that out of the three students who took part in all six test/retest cycles, two showed a decrease in performance in retest.

One of the students to complete all six cycles had both the lowest average test and retest scores among all participants. It would be hard to argue using this participant's data in isolation that feedback interventions can have greater effects with lower achieving students (Black & Wiliam 1989) than whether learners need to have an acceptable level of subject familiarity (Bereiter and Scardamalia 1993 in Sadler 2013) before such feedback interventions can have a significant effect or that the marks given (that the students could see) for the original test act as a negative confirmation (Ecclestone 2003) to the participant, confirming what they think of their ability and dissuading them from remedial work to improve. A wider issue, not just relevant to the participants mentioned here, is the freedom of learners to accept or reject any feedback that is given (Kulhavy 1977, in Hattie & Timperley 2007) or even modify it, and that the giving of feedback itself is not necessarily a precursor, as will be seen later in this chapter, to significant action upon it.

Out of the three participants whose test averages were lower (70%) than the pass mark for the external exams on which the course was finally assessed, one student (the student mentioned above) showed decreased average performance on retest, one student showed no overall change on retest and one student showed an increase on performance on retest, taking them from over the 70% threshold with an 8% increase (table 4), that gives us a mixed picture with regards to the effects of the feedback intervention on lower achieving participants in particular.

Table 4: Individual test/retest cycle scores and percentage increase or decrease

Unique ID	Test/Retest cycles	Avge Test	Avge Retest	% Incr/Decr
1	5	92.00	96.00	+4.00
2	6	61.67	60.00	-1.67
3	4	72.50	75.00	+2.50
4	4	100.00	100.00	0.00
5	5	80.00	96.00	+16.00
6	4	95.00	90.00	-5.00
7	4	85.00	76.25	-8.75
8	4	70.00	70.00	0.00
9	3	90.00	76.67	-13.33
10	6	90.00	88.33	-1.67
11	3	93.33	93.33	0.00
12	5	65.00	73.00	+8.00
13	5	90.00	100.00	+10.00
14	2	75.00	90.00	+15.00
16	3	83.33	90.00	+6.67
17	3	66.67	66.67	0.00
18	3	90.00	80.00	-10.00
19	6	97.00	99.00	+2.00
22	5	80.00	84.00	+4.00
23	2	85.00	85.00	0.00

As can be seen from table 4, a total of nine participants increased their average scores on retests, with improvements ranging from +2% to +16%, with most improvements being in the range of +2 to +8% and three participants improving by +10% or greater. Five participants showed no overall change in their average test/retest scores, although as mentioned earlier three of those showed some increase or decrease within individual test/retest cycles and two participants having identical scores in each cycle. Note that one participant who showed no overall change scored full marks throughout all four test/retest cycles completed. Six participants showed an overall decrease in their retest scores compared with original

test scores. Two participants decreased by -10% or greater with other average decreases ranging from -1.67% to -8.75%.

Use of Feedback

An important part of interpreting the data presented above was what participants did with any feedback that they were given, and so participants were asked how they noted any feedback received with four options: making notes on paper of the feedback given, printing off feedback and quiz answers, taking a picture of the computer screen showing feedback received, or doing nothing with the feedback, i.e. no notes, pictures or printouts. Participants could select one or more methods used during the research study and were not limited to selecting one method. This question was not related to what participants did with feedback subsequently with regards to revision and preparation for the retest (which was asked in a subsequent question), but rather how they noted (i.e. stored) information regarding any feedback received.

Table 5: Feedback storage methods used by students

	Students with average increase on retest	**Students with average no overall change**	**Students with average decrease on retest**
Make paper notes on feedback given on VLE	4	3	4
Print off feedback from VLE	5	0	1
Take a picture of feedback from VLE	1	1	1
No notes taken of feedback	2	0	2

Table 5 shows the four methods used for storing feedback. The first method, making notes on paper, showed relatively even distribution although only just over half of students used this method. Only three students took pictures of their computer screen, and these were also evenly distributed, and those who made no notes, showed even distribution between increasing or decreasing performance on retest. However, a fourth method (printing feedback from the VLE whilst still in class) showed a clear apparent correlation with participants whose scores increased on retest. 5 out of 6 students who used this method of storing feedback, and 5 out of 9 students who showed an overall increase in retest scores, printed feedback from the VLE as part of their feedback storage and revision progress. However, the single participant who decreased on retest after printing off feedback from the VLE was the participant with the lowest decreasing average score (1.67%) on retest, took part in all six test/retest cycles and had no scores below 70% for any tests. The relatively

small sample size of this study means that too much must not be made of this, but it is an area of interest and potential further research about what it is about printing off such feedback that may improve student performance. It is worth considering that the proactive nature of the printing out of feedback (the printer was in a shared hallway in a corridor and required students to get up and leave the room and log in to the printer) may suggest something about the proactive nature of the participants in then using that feedback to improve their learning.

Participants were then asked what methods they used in the six days between the test and the retest for each cycle to revise the topic to prepare themselves for the subsequent retest. Seven options were provided based upon the common options available and used as part of the course structure and content, plus the option to indicate that no revision took place. The first three methods involved re-reading materials that already existed, including materials in the set text books for the course, and re-reading notes that the participants had themselves made or the presentations (using Microsoft PowerPoint software) that were available on the VLE and provided by the course tutor. The next two methods involved progressing through e-learning modules or answering short quizzes on the website of the qualification's awarding body using their student member login. The final method involved re-taking any tutor-provided quizzes separate from the research study tests/retests that were available on the VLE. Participants were again able to choose more than one method.

Table 6: Revision methods used by participants prior to retest

	Students with average increase on retest	**Students with average no overall change on retest**	**Students with average decrease on retest**
Re-read course materials in course books	6	3	6
Re-read own notes	5	3	5
Re-read presentations on VLE	3	1	4
Online e-learning modules provided by awarding body	4	2	3
Online tests provided by awarding body	5	2	3
Chapter activities from course books	5	2	5
Tutor-provided activities on VLE	3	0	3
Nothing	2	0	0

One of the most interesting pieces of data to be extracted from this piece of information is that the two participants who indicated that they did no revision were both participants who increased their average score on retest, one by the second highest margin of +15% (which showed an average test/retest score of 75%/90%) and one with the second highest test/retest (97%/99%) score. However, it should also be noted that the first participant indicated here only participated in two test/retest cycles due to absence. This and the small number of students, (two out of twenty, or 10%) who indicated that they did no revision between test and retest means that we cannot produce any firm conclusions on this particular data, other than to note its interest.

Four of the methods (re-reading course books, re-reading own notes, online e-learning modules and chapter activities from course books) showed little or no significant data that might explain an increase or decrease in performance. Re-reading tutor-provided presentations on the VLE involved more students (four) decreasing on retest than increasing (three) although it must be remembered that students were using a variety of methods. Numerous authors have suggested (Hattikudur & Postle 2011, Roediger et al. 2011, Butler & Roediger 2007, Carpenter 2012) that the testing effect, whereby students are tested on materials to strengthen retrieval practice, has a greater effect than students merely re-reading course materials.

When participants took online tests provided by the awarding body or took part in tutor-provided activities on the VLE, 50% (five out of ten) of those students showed an average increase on retest, a slightly higher number than the overall group average (9 out of 20, or 45%). Although not showing great statistical significance, it may be of importance that both of these methods involve answering short-answer and multiple-choice questions in ways that are not dissimilar to the style of the questions asked of participants in the research study. This small but interesting raise in performance may suggest the significance of retrieval practice, practice in retrieving the knowledge that has been learned, as a crucial factor in improving performance in that it improves retrieval strength, the ability of the learner to retrieve information (Bjork & Bjork 1992, 2006).

Having considered how participants stored their feedback for later use and what methods they used to act on feedback that had been given, it was also important to consider how long students spent with that feedback in terms of reflection and preparation for the retest. Although time spent is not necessarily an indicator of the effectiveness of revision as raw indicator, it can be an indicator of effort.

Table 7: Time spent by participants with feedback/revision prior to retest

	Students with average increase on retest	**Students with average no overall change on retest**	**Students with average decrease on retest**
0-15 minutes	1	0	2
15-30 minutes	2	2	0
30-45 minutes	1	2	2
45-60 minutes	2	0	1
60+ minutes	3	1	1

It is noticeable from the information provided in table 7 that five out of nine students (55.56%) who showed increased scores on retest claimed to have spent more than forty-five minutes in revision processing the feedback and with further study, compared with two out of six (33.33%) of students who showed decreased performance in retests. Thirteen out of the twenty students said that they had spent at least half an hour after each test processing feedback and preparing for retest. The student who increased scores on retest with fewer than fifteen minutes of feedback-led revision was the participant with the highest test scores (97%) other than the student who achieved perfect 100% scores on all tests & retests, the latter stating that they did forty-five to sixty minutes of feedback-led revision.

There is some correlation in these figures between amount of time spent in feedback-led revision and retest performance, but as can be seen two participants spent greater than forty-five minutes on revision and decreased performance on the subsequent retest and one participant showing no overall change. Notably, all students who had spent more than forty-five minutes and had shown an increase in retest performance stated that they had also printed off feedback from the VLE to aid with their revision. These figures indicate some patterns between methods of feedback storage, revision methods and time spent revising on the individual level, but some exceptions and sample sizes mean that firm, generalisable conclusions cannot be made.

A further issue to be considered when interpreting these figures is the individual study skills of participants and their ability to process the feedback and revise successfully from it, in a course where there was no time to significantly help participants develop or refresh their study skills. The varied ages of the participants and varied previous exposures to adult education meant that some had studied recently whereas some had not studied for many years. It could also be that those who studied recently might have inefficient study skills whereas those who studied some decades ago had efficient study skills and study patterns, so presumptions cannot and should not be made on age alone. The researcher must acknowledge that the ability or lack of ability of participants to process the feedback that they have been given and to revise effectively, or have the time to revise effectively because of adult and family responsibilities, will most likely have an effect on the final data gathered.

Working Patterns

As has been discussed within the literature review, much of the research into testing and feedback has predominantly been on younger students with a large focus on opportunity or convenience samples of school, undergraduate or postgraduate students. A distinctive feature of the cohort of adult students for this research study was that many of them worked during the day and then came to class in the evening, and so an important part of analysing the data for this study was the potential effects of working patterns on performance.

When considering the effects of working patterns and employment on performance, it is also important to consider the effects of the actual work undertaken by participants. Some participants worked in finance and accounting, although in different areas from bookkeeping, and therefore would have some domain knowledge that they could bring into the learning environment that may improve their performance, particularly when applying pre-existing conceptual frameworks and knowledge to the correction of errors and the implementation of feedback. Conversely, those participants who did not work in accounting or finance and had no such pre-existing domain knowledge perhaps beyond knowledge on banking and their own personal finances, might not have such conceptual frameworks to draw upon. Therefore, when attempting to analyse and interpret test/retest results against working patterns, it is useful to bear in mind the employment situation and any known specifics of participants in this regard.

Table 7: Working patterns of participants

	Students with average increase on retest	**Students with average no overall change on retest**	**Students with average decrease on retest**
No hours worked	2	0	1
Up to 2 hours	0	0	0
2-4 hours	0	0	1
4-6 hours	0	0	0
6-8 hours	4	1	1
8+ hours	2	4	3
No information given	1	0	0

It had been observed prior to the commencement of the study, and many participants had commented on it themselves, that they often arrived tired to the evening session, having come directly from their places of employment and in many cases without opportunity to eat or drink significantly before the class started. An analysis of information provided by participants with regards to their working patterns on the days that they attended the learning provider has provided information that warrants further research. However, again it must be noted that this information relates to relatively small sample sizes and that any findings based upon this information are conditional. It is of interest that the four students showing the highest averages in performance on retest (see table 8 below) all worked between 6-8 hours on the days of the test/retest cycles. The working patterns of participants were consistent, in that the working patterns all patterns were consistent across the two days in which they attended the learning provider for study.

Table 8: Working patterns of individual participants, including individual average test results

Unique ID	Incr/Decr %	Avge Test	Avge Retest	Tuesday	Wednesday
5	+16	80	96	6-8 hours	6-8 hours
14	+15	75	90	6-8 hours	6-8 hours
13	+10	90	100	6-8 hours	6-8 hours
12	+8	65	73	6-8 hours	6-8 hours
16	+7	83	90	No	No
22	+4	80	84	No info	No info
1	+4	92	96	8+ hours	8+ hours
3	+3	72	75	No	No
19	+2	97	99	8+ hours	8+ hours
4	0	100	100	8+ hours	8+ hours
8	0	70	70	6-8 hours	6-8 hours
11	0	93	93	8+ hours	8+ hours
17	0	67	67	8+ hours	8+ hours
23	0	85	85	8+ hours	8+ hours
2	-2	62	60	8+ hours	8+ hours
10	-4	90	86	2-4 hours	2-4 hours
6	-5	95	90	6-8 hours	6-8 hours
7	-9	85	76	8+ hours	8+ hours
18	-10	90	80	No	No
9	-13	90	77	8+ hours	8+ hours

Of interest are those participants who either did not work at all in the day leading up to the evening session, or who worked more than six hours during that day. Two students who did not work during the day showed an average increased (+7% and +2.5%) performance on retest (whilst one student who did not work during the day showed an average decrease (-10%) during retest. A potential reason for the students who showed an increase on retest without working during the day could be that they arrived at the evening sessions in a fresher state of mind, having not been

tired out during the day by long hours. However, that does not then explain the one student who showed an average decrease in performance on retest of -10% having not worked during the day. A possible reason for this might be that the student has not been sufficiently intellectually challenged during the day and so has struggled to get up to speed mentally once the class has begun. There are of course may be other individual factors for those students, information for which is beyond the reaches of this study that explain the differences in performance, for example generalised or test-based anxiety. Anecdotally, one student who did not work often stated that they had been studying during the day prior to coming to class, and this student showed a mild (+2.5%) average increase in performance on retest.

As mentioned above the highest correlation between working patterns and performance was between those who worked six-eight hours (four out of six students showing an average increase in retest performance) and eight+ hours work a day (seven out of nine students showing either no change in average retest performance or a decrease in average retest performance). For the top four performers, three had taken part in five test/retest cycles and one (with the second highest average increase of +15%) had only taken part on two test/retest cycles. The participants working six-eight hours whose average retest scores either remained the same both took part in four test/retest cycles, with the student whose retest performance decreased on average having a higher average test performance of 95%, decreasing on retest to 90%.

Working more than eight hours on the days on which participants attended the learning provider showed a stronger inclination towards either no overall change in

average performance (four students) or a decrease in average performance (three students), although two students also showed an average increase in performance having worked for over eight hours.

One of the two students who increased on average test performance having worked more than eight hours already worked in accountancy albeit in a different area, and the relevant domain knowledge already gathered by this participant may have given them an advantage, or negated the disadvantage of longer working hours.

One participant who showed no overall change on retest performance having worked eight or more hours prior to test and retest was the one participant who scored 100% on every test and retest. This student did not work in a related accountant or financial field, and was at the time of the research study a forklift truck driver, and so it seems unlikely that any specific domain knowledge was carried over between the two disciplines. Two of the other three participants who showed no overall change in retest performance worked in related accounting fields, which again might have helped performance in this regard or work to limit any negative effects of longer working hours.

Out of the three participants whose performance in retests decreased after working eight or more hours in a day (-9% decrease) did work in a closely-related accounting field, but did much of their work using computer systems rather than manual bookkeeping methods and knowledge, which was what was predominantly tested. Predominantly, modern bookkeeping software is designed for ease of use

and often hides the more complicated elements of bookkeeping from the user unless the user should go looking for it. One participant, who showed the greatest overall decrease of -13.33% worked in a school-based environment with some exposure to finance but no exposure to accounting methods and practices, with the other participant showing a decrease in retest performance working in blue-collar industry as a fitter.

The significance, if any, of working patterns on performance in students in an adult evening class is worth further examination. This particular cohort of students regularly (approximately every six weeks) took exams as summative assessments for the course which then led to graded (pass, merit, distinction) final marks for the qualification. Comparison might be useful between performance of an evening class and a daytime class in order to see whether any lessons can be learned in this regard and what can be done to improve performance.

Perception of Improvement

Finally, participants were asked their perceptions of the feedback, in terms of their perceived impact on their scores on retest. This used a five-point Likert scale (see table 9 overleaf).

Table 9: Participant perception of effect of feedback on retest performance

	Students showing average increase on retest	**Students showing average no overall change on retest**	**Students showing average decrease on retest**
Considerable positive impact	3	1	2
Some positive impact	6	4	3
No impact	0	0	0
Some negative impact	0	0	1
Considerable negative impact	0	0	0

All students except one suggested that the feedback intervention had at least some positive impact on their retest scores. All five students who had no overall change on test and retest scores and five out of the six students who on average decreased on their retest scores stated they believed that their retest scores were higher on average than their test scores. Participants had access to their test scores on the VLE until the end of the class that evening rather than through the length of the study. It may be that the feedback intervention, in nineteen out of twenty (95%) cases made the participants feel more confident about their learning, even if this was not borne out for the most part in the actual retest scores. It is also possible that the participants felt a placebo effect, in that they believed that the feedback intervention worked, or that the feedback intervention and the retrieval practice had a longer-term effect on the learning of the participants, given that the question with regards to perception of the feedback intervention on retest performance was asked after the end of the research study and just after an exam taken by the participants.

Conclusion

An analysis of the data suggests that there some mild correlations between different elements of the data, and that an increased score on retest may be linked to effective storage of feedback information, effective revision techniques, working patterns and a sufficient amount of time spent on revision prior to retest, but with a few exceptions. The data and the subsequent analysis highlight some of the inherent complexities of adult learning and adult performance in the context of weekday evening classes.

Conclusions & Recommendations

Introduction

In outlining conclusions as a result of data gathered in this study and making recommendations for the future, it is important to bear in mind the context behind this study and dissertation. This dissertation is set within the context of a qualification in professional practice in education. This means that it will be appropriate not just to consider the issue from an academic perspective, but also in terms of what this might mean for professional practice, for myself and for others.

Conclusions

This study began with the theoretical assumption that feedback and testing could improve student performance and help students to close the gap between what they know and what they should know at any given point in their studies. It was suggested in the literature review that although there is considerable evidence that shows that feedback and testing can improve learning, that there are caveats in that it does not happen in all cases. The results of this study have backed up this initial view of feedback and testing, in that the results of the pre-written feedback in improving learner performance were mixed. Overall, nine students showed some average improvement, five students showed no average overall change and six

students showed an average decrease in performance. There seem to have been several reasons for this.

Firstly, students who increased their scores on retest had lower average test scores than those whose average scores decreased on retest. This might have been because students who scored higher on their original retest either had less space to move upwards in terms of marks given, or that the relatively high scores in the first test made them less inclined to revise thoroughly or effectively than those students who performed lower on the original test who had more room for improvement.

Secondly, there was considerable variation between students on how they dealt with feedback that they received and how they factored that feedback into subsequent revision. Those students who printed off their feedback, spent a considerable amount of time on revision and used a variety of methods including testing as part of their revision, were more likely to improve on retest than those who spent shorter periods of time or used inefficient methods of revision such as re-reading notes and course materials. It is possible that there was a disconnect in many cases where the feedback was given by the tutor but not fully received by the students, in that they did not appreciate fully what to do with the feedback that they were given or they did not sufficiently understand it. A tutor may presume that once they have given feedback to a student, either verbally or in writing, that their job is done. However, through either miscommunication, misunderstanding, or an unwillingness to work with feedback, this is not always the case.

Thirdly, working hours patterns seemed to have some impact on student performance. The students whose performance increased the most worked between six and eight hours each day, whilst those who worked more than eight hours per day were more likely to have no average overall change in performance or decrease in performance on the retest, although those six students whose performance was lower on retest exhibited the widest range of working hours patterns.

It must be remembered that these tentative conclusions are drawn from a relatively brief period of research with a small cohort of students compared with some other recent research, and that such conclusions cannot be used to draw general conclusions about feedback and testing in education, and that they raise questions as much as they provide answers, and as such the recommendations include a request for more information and study.

Recommendations

The first recommendation is that the research study be repeated at some point in the near future, with some small changes. These changes would involve giving students more guidance at the beginning of the study with regards to what to do with feedback to hopefully improve the result. Students could be told about the benefits of testing and quizzing versus merely re-reading materials, and encouraged to take online tests more and complete more question/answer tasks as a revision tool prior to retest. Results of the subsequent study can then be compared with this study to see if there are any differences. Students would once again be asked

questions with regards to their working patterns and how they noted feedback and used that in their revision, to see if it fills in gaps that this study has left open. It will also be necessary within this to do some work with students on study skills, to ensure that students are giving themselves the greatest opportunity to succeed at the highest possible level of performance. It is important for the tutor to work on ensuring that the student not only receives but also understands the feedback sufficiently to process and act upon it.

Another option is to repeat the research study, but this time with a daytime class of accounting students studying the same course, and note any differences in performance from the evening classes. Most students in daytime classes will, it is suggested from experience, either be working part-time or working around the course, perhaps on day release from their place of work. These factors may bring in another dynamic to help explain the performance of adult students. It is also an option to repeat the study, but with set of adult students with the same training provider studying different courses, for example leadership and management or personnel and development. Results from such students on different courses might bring more depth and breadth to an understanding of feedback and testing with adult students.

Secondly, it is suggested that more work needs to be done in terms of the particularities of adult learners and how they can be best supported through feedback to close the gaps in their learning as much as possible. More research could be undertaken that helps to find the ways in which adult learners can best encode and process feedback and use it to their best advantage. Many adult

students may be very time-poor in terms of what time they have available to study and may need more directing, even in spite of the apparent self-directed nature of many adults and the experiences that they bring to the learning environment. Many adult professional courses, such as accounting, have relatively high drop-out rates particularly because of performance early in the course demotivating some students, and more effective feedback as part of assessment for learning could help to lower these figures.

More research and study of the most effective forms of feedback and testing in adult students would help to enrich the research literature as well as providing tutors and teachers of adults as practitioners with valuable information that will help their students maximise their chances of achievement.

Action Plan

Goal	Purpose	Resources	Timeframe	Deadline
Feedback to colleagues in business & other departments	Inform colleagues (particularly those teaching adults) of findings	MEd dissertation & research materials MEd presentation PowerPoint College VLE	0-3 months	December 2017
Repeat study with next cohort of level 2 accounting (evening) students)	To check results against initial study, test further hypothesis against additional information re: further help for students on processing feedback and doing revision	Already available resources from initial study Additional resources to include: • Resources on working with feedback • Resources on effective revision • Resources on study skills	3-6 months (from January 2018)	July 2018
Seek publication of studies 1 & 2	To add to the literature on feedback and testing for adults	MEd dissertation & research materials Results/information from previous study List of appropriate journals to submit (education and/or accounting)	12 months (from submission to publication, approximate)	July 2019 *(1 year after initial study)*

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (eds.), The psychology of learning and motivation: Advances in research and theory (vol. 2, pp. 89- 195). New York: Academic Press.

BERA (2011) 'Ethical guidelines for educational research', accessed 4th May 2017 Available at: *https://www.bera.ac.uk/wp-content/uploads/2014/02/BERA-Ethical-Guidelines-2011.pdf*

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes.* (Vol. 2, pp. 35– 67). Hillsdale, NJ: Erlbaum.

Bjork, R. A., & Bjork, E. L. (2006). Optimizing treatment and instruction: Implications of a new theory of disuse. In L-G. Nilsson and N. Ohta (Eds.), *Memory and society: Psychological perspectives.* London: Routledge

Black, P., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. (2002) *Working inside the black box: assessment for learning in the classroom* London: King's College

Black, P. & Wiliam, D. (1998) *Inside the Black Box: Raising Standards Through Classroom Assessment* London: King's College

Butler, A.C. & Roediger, H.L. (2007) 'Testing improves long-term retention in a simulated classroom setting' in *European Journal of Cognitive Psychology* vol. 19 issue 4, pp514-527

Butterfield, B. & Metcalfe, J. (2006) 'The correction of errors committed with high confidence, in *Metacognition Learning* vol. 1 pp. 69-84

Byrne, M., Flood, B. & Griffin, J. (2014) 'Measuring the academic self-efficacy of first-year accounting students' in *Accounting Education* vol. 23 issue 5, pp.407-423

Carpenter, S. (2012) 'Testing enhances the transfer of learning' in *Current Directions in Psychological Science* vol. XX(X) pp.1-5

Carrier, M. & Pashler, H. (1992) 'The influence of retrieval on retention' in *Memory & Cognition* vol. 20 no. 6 pp.633-642

Churches, R. & Dommett, E. (2016) *Teacher-led research: designing and implementing randomised controlled trials and other forms of experimental research* Bancyfelin: Crown House

Cohen, L., Manion, L. & Morrison, K. (2013) *Research Methods in Education, 7th ed.* London: Routledge

De Brucyckere,P., Kirschner, P.A., Hulshof, C.D. (2015) *Urban Myths about Learning and Education* Academic Press: London

Ecclestone, K. (2003) *Understanding assessment and qualifications in post-compulsory education: principles, politics and practice* Leicester: NIACE

Ecclestone, K., Davies, J., Derrick, J. & Gawn, J. (2010) *Transforming Formative Assessment in Lifelong Learning* Maidenhead: Open University Press

Economist , The (2017) 'Technology is transforming what happens when a child goes to school', accessed 21st July 2017 Available at: *https://www.economist.com/news/briefing/21725285-reformers-are-using-new-software-personalise-learning-technology-transforming-what-happens*

Eisenkramer, R.E., Jaeger, A. & Stein, L.M. (2013) "A systematic review of the test effect in learning" in *Paideia vol. 56 no.23* pp. 397-406

Gibbs, J.C. & Taylor, J.D. (2016) 'Comparing student self-assessment to individualized instructor feedback' in *Active Learning in Higher Education* vol. 17 no. 2 pp. 111-123

Gipps, C.V. (1994) *Beyond testing: towards a theory of educational assessment* London: Falmer Press

Guasch,T., Espasa, A., Alvarez, I.B., Kirschner, P.A. (2013) 'Effects of feedback on collaborative writing in an online learning environment' in *Distance Education,* volume 34 issue 3, pp.324-338

Hattie, J. (2012) *Visible learning for teachers: maximizing impact on learning* London: Routledge

Hattie, J. & Timperley, H. (2007) 'The Power of Feedback' in *Review of Educational Research*, vol. 77 no.1, pp. 81-112

Hattikudur, S. & Postle, B.R. (2011) 'Effects of test-enhanced learning in a cognitive psychology course' in *Journal of Behavioural and Neuroscience Research* vol. 9 issue 2, pp.151-157

Hinze, S.R. & Rapp, D.N. (2014) 'Retrieval (sometimes) enhances learning: performance pressure reduces the benefits of retrieval practice' in *Applied Cognitive Psychology,* vol. 28 pp.598-606

Innes, A. (2009) *Dementia studies: a social science perspective* London: Sage

Jarvis, P. (1995) *Adult & continuing education: theory and practice* 2nd ed. London: Routledge

Jarvis, P. (2014) 'From adult education to lifelong learning and beyond' in *Comparative Education* vol. 50 issue 1 pp.45-57

Kang, S.H.K., McDermott, K.B., and Roediger, H.L. (2007) "Test format and corrective feedback modify the effect of testing on long-term retention" in *European Journal of Cognitive Psychology* vol. 19 no. 4 pp.528-558

Kluger, A.N. & DeNisi, A. (1996) 'The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory' in *Psychological Bulletin* vol. 119 no.2, pp.254-284

Knowles, M. S. (1980) *The modern practice of education: from pedagogy to andragogy* 2nd. Ed New York: Cambridge

Knowles, M.S., Holton III, E.F. & Swanson, R.A. (2014) *The Adult Learner* 8th ed. London: Taylor & Francis

Lamb, G. (2016) 'Raising learner participation in a virtual learning environment through an online forum' in *Through the Looking Glass: Reflective Research in Post Compulsory Education* volume 11 pp.95-102

McDowall, T., Jackling, B. & Natoli, R. (2015) 'Relationships between vocational interests and learning approaches to advance the quality of student learning in accounting' in *Accounting Education* vol.24 no.6 pp498-513

Metcalfe, J. (2017)'Learning from errors' in *Annual Review of Psychology* vol. 68 issue 6, pp.6.1-6.25

Narciss, S. (2008). 'Feedback strategies for interactive learning tasks'(pp.125-144) In J. M. Spector, M. D. Merrill, J. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology (3rd ed.)* Mahwah, NJ: Lawrence Erlbaum.

Panhoon, S. & Wongwanich, S. (2014) 'An analysis of teacher feedback for improving teaching quality in primary schools' in *Procedia - Social and Behavioral Sciences* Issue 116 pp.4124-4130

Roediger, H.L. & Karpicke, J.D. (2006) 'Test-enhanced learning: taking memory tests improves long-term retention' in *Psychological Science* vol. 17 no.3 pp.249-255

Roediger, H.L., Agarwal, P.K., McDaniel, M.A. & McDermott, K.B. (2011) 'Test-enhanced learning in the classroom: long-term improvements from quizzing' in *Journal of Experimental Psychology: Applied* vol. 17 no.4 pp.382-395

Rogers, A. (2003) *What is the difference? A new critique of adult learning and teaching* Leicester: NIACE

Sadler, D.R. (1989)'Formative assessment and the design of instructional systems' in *Instructional Science* vol. 18, pp.119-144

Sadler, D. R. (2013) 'Opening up feedback: Teaching learners to see'. In Merry, S., Price, M., Carless, D., & Taras, M. (Eds.) *Reconceptualising Feedback in Higher Education: developing dialogue with students*. (Ch. 5, 54-63). London: Routledge.

Smith, M.A. & Karpicke, J.D. (2014) 'Retrieval practice with short-answer, multiple-choice, and hybrid tests' in *Memory*, vol. 22 no.7 pp.784-802

Stobart, G. (2006) 'The validity of formative assessment' in Gardner, J. (2006) *Assessment and Learning* pp.133-146 London: Sage

Turner, M. & Baskerville, R. (2013) 'The experience of deep learning by accounting students 'in *Accounting Education* vol. 22 no.6, pp.582-604

Usher, R., Bryant, I. & Johnston, R. (1997) *Adult education and the postmodern challenge: learning beyond the limits* London: Routledge

Appendix A – Ethics Proposal

University of Central Lancashire

Research Ethics - CELT

Application Form

Undergraduate & Postgraduate Taught Students

Name of student	**Graeme Lamb**	Email	**GWLamb@uclan.ac.uk**
Name of personal tutor/supervisor	**Paul Doherty**	Email	**PWDoherty@uclan.ac.uk**
Degree registered for (please tick)	**MEd**		
Title of proposed research	**The effect of pre-written feedback for online tests in improving learner performance**		
Brief summary of project aims	To research the effects of pre-written feedback in improving learner performance in online tests		
Brief summary of project methodology	Adult learners will complete a weekly online quiz for six to eight weeks, asking questions based on the learning for that week. Pre-written feedback will tell them the areas to focus on for revision, and they will then be retested to measure improvements.		

All materials submitted to the CELT Research Ethics Committee will be treated confidentially.

The checklist below should be completed with the assistance of your dissertation supervisor. This checklist will identify whether a project requires an application for ethics approval, and whether it needs to be submitted directly to the CELT Research Committee.

Please refer to the University Code of Conduct for research (above) Dissertation supervisors are responsible for exercising appropriate professional judgement in

undertaking this review and evaluating the research proposal according to the criteria laid down in the checklist.

Please answer the following questions:

1	Does the study involve participants who are unable to give their informed consent (e.g. children, people with severe learning disabilities, unconscious patients etc.) or who may not be able to give valid consent (e.g. people experiencing mental health difficulties)?	**NO**
2	Does the project raise issues involving the potential abuse or misuse of power and authority which might compromise the validity of participants consent (e.g. relationships of line management or training)?	**NO**
3	Is there any potential risk arising from the project of physical, social, emotional or psychological harm or distress to the researchers, participants or audience?	**NO**
4.	Does the project involve a potential risk of causing shock, offence or outrage to researchers, participants, the audience or public?	**NO**
5	Does the project involve researchers and/or participants in the potential disclosure of any information relating to illegal activities; the observation of illegal activities; or the possession, viewing or storage of any material (whether in hard copy or electronic format) which may be illegal?	**NO**
6	Will the deception of participants be necessary during the study?	**NO**
7	Will the study involve invasion of privacy or access to confidential information about people without their permission (if 'YES' see note below)?	**NO**
8	Will the study involve any external organisation for which separate and specific ethics clearance is required (such as the NHS; any criminal justice agencies including the Police, Crown Prosecution Service, Prison Service, Probation Service or successor organisations)?	**NO**
9	After completing the Data Protection compliance checklist are there any data protection compliance problems?	**NO**

Appendix B – Consent Form

Information sheet for participants taking part in research project

Study title

The effect of pre-written feedback for online tests in improving learner performance

Invitation to participate

You are being invited to participate in this study. However before doing so it is important to fully understand the reasoning behind the research. I would strongly advise that you read the following information and discuss with others if you feel the need to. If you are unclear at any point and require further clarification please do not hesitate to contact me at the contact details below. Please do take the time to consider your involvement in this study before consenting.

Thank you for taking your time to read this information sheet and please retain it for your records, should you decide to partake in the study.

What is the purpose of the study?

The aim of this study is to test whether pre-written feedback as a response to answers given in online quizzes in a virtual learning environment (Virtual Campus) improves performance amongst learners.

Why have I been asked to take part?

Your participation in this study will help provide valuable data as to the effects on pre-written feedback in response to questions on an online quiz, hopefully informing educational practitioners as to its benefit as part of teaching, learning & assessment

Do I have to take part?

You are under **no** obligation to take part in this study. If you do decide to participate you will be asked to retain this information sheet as well as sign a consent form. Your decision to participate does not bind you to the study and you are free to withdraw at any point without reason.

What do I have to do if I decide to take part?

As part of this study you will be invited to take part in a weekly online quiz to be taken at the end of the Wednesday night class on a specific topic and then a subsequent test on the same material in class during the next (Tuesday) session. This would be expected to take no more than 10-20 minutes of class time per week for up to 6 weeks.

What are the benefits and risks of taking part?

The intended benefit is that the pre-written feedback given will help to focus your revision on specific areas of improvement rather than just generic revision. The only potential risk is the time invested should there be no significant results indicated the effectiveness of the study.

What if I have a complaint about the study?

If you are unhappy at any point of the study and wish to raise this concern then please contact the researcher Graeme Lamb Graeme.lamb@wmc.ac.uk or the course leader Dr Candice Satchwell on CSatchwell@uclan.ac.uk

Will my taking part in this study be kept confidential?

The lead researcher, Graeme Lamb, is the only person that will have access to the information gathered once it has been collated and anonymised. Prior to that, responses to VLE questions will be available to any tutor that has tutor access to the relevant pages, as per the normal operation of the VLE at college. Confidentiality will be maintained as far as possible at all times, and no personally identifiable information will be included in the final published results.

As this study will be conducted through the University of Central Lancashire the researchers will be required to adhere to the University regulations. The regulations state that the information collected will be stored for a maximum of 5 year and I will be keeping the results for 3 years. Any electronic information with regards to the study will be kept on password-protected systems at all times.

What will happen to the results of the research study?

The results of the study will be published as an MEd thesis by UCLAN. The author of the research may seek further publication in an educational academic journal if the results are considered significant enough to do so.

Who is organising and funding the research?

This research is organised by the MEd student, Graeme Lamb, under the supervision of staff at the University of Central Lancashire. There is no funding involved in this research project

Who has given permission for the study to go ahead?

The Ethics Committee within CELT based at the University of Central Lancashire has approved the research to be completed.

Contact for further information

If you require further information, please contact me, Graeme Lamb Graeme.lamb@wmc.ac.uk or the course leader Dr Candice Satchwell on CSatchwell@uclan.ac.uk

Thank you for considering taking part in this research.

Consent Form for Participants:

The effect of pre-written feedback for online tests in improving learner performance

Please acknowledge that you agree with the following statements:

I understand the purpose of this research and the requirement to regularly take part in online activities on the college VLE (Virtual Campus)

Please tick to confirm □

I understand that my participation is voluntary and I can opt out at any point. I understand that I do not have to answer all the questions if I do not feel comfortable answering. I understand I do not have to take part in the activity.

Please tick to confirm □

I have been offered the opportunity to ask any questions relating to this research and they have been answered to my satisfaction.

Please tick to confirm □

I understand that I am free to withdraw at any time, without giving any reason and my wishes will be respected.

Please tick to confirm □

I understand that any data about me will remain confidential and my identity will be anonymised unless I disclose information which suggests that I or someone else may be at risk of serious harm.

Please tick to confirm □

I agree that I have read and understood the information sheet.

Please tick to confirm □

I am happy to take part in this research.

Please tick to confirm □

Name: _____

Signature: _____

Date: _____

Appendix C – Sample Questions and Feedback

Question: *When completing a journal entry, the correct format is list the debit entry before the credit entry, True or False?*
Answer: *true* **Feedback for correct answer:** *you have correctly identified that the debit entry is listed before the credit entry. This is done so that accountants and bookkeepers keep records in an identical manner* **Feedback for incorrect answer:** *incorrect, the debit entry should always be listed before the credit entry. This is done so that accountants and bookkeepers keep records in an identical manner. Your revision should focus on the rules of the journal and making journal entries*
Question: *When recording a credit sale, the correct format is: credit Sales & Sales Ledger Control, debit VAT. True or false?*
Answer: *false* **Feedback for correct answer:** *correct, you have identified that this is not the correct combination of entries, which should be: credit Sales & VAT, debit Sales Ledger Control Account. The amount of the Sales Ledger Control Account will be the sum of the other two (Sales & VAT)* **Feedback for incorrect answer:** *incorrect, the rule is that the sales (which is the net amount) and the VAT should be on the same side as each other (credit) and will add up to the amount on the debit side (Sales Ledger Control Account). Your revision should focus on making initial entries for credit sales and credit purchases*
Question: You are preparing an invoice for a longstanding customer who receives 10% trade discount. The net amount on the invoice before the discount is applied is £1,200, what will be the total gross payable once the discount is applied?
Answer: *£1,296 (£1,080 + VAT @ 20% of £216)* **Feedback for correct answer:** *correct, you have discounted the invoice of £1,200 to reduce the net to £1,080 and then you have correctly applied VAT @ 20% of £216 to make the gross total £1,296* **Feedback for incorrect answer:** *incorrect, you first need to remove 10% trade discount from the net amount (£1,200 – 10% = £1,080) and then add VAT @ 20% (£1,080 x 20% = £216), £1,080 + £216 = £1,296. Focus your revision on trade discounts and VAT calculations*

Question: *From the figures provided, calculate the total wages expense to the employer*

Gross wages: £25,000
Tax: £3,750
Employers NI: £1,200
Employee NI: £1,200
Employee Pension Contribution: £800
Employer Pension Contribution: £800

Answer: £27,000 (£25,000 + £1,200+£800)
Feedback for correct answer: *correct, you have identified that the total wages expense is the gross wages (how much employees have earned) plus any employer-only (NI & pension) contributions*
Feedback for incorrect answer: *incorrect, the total wages expense to the employer is the gross wages (how much employees have earned (£25,000) plus any employer-only (NI £1,200 & pension £800) contributions> in your revision focus on employer contributions and total employer wage costs*

Question: *From the figures provided, calculate the liability to HMRC (HM Revenue & Customs)*

Gross wages: £25,000
Tax: £3,750
Employers NI: £1,200
Employee NI: £1,200
Employee Pension Contribution: £800
Employer Pension Contribution: £800

Answer: *£5,750 (£3,750 + £1,200 + £1,200)*
Feedback for correct answer: *correct, you have identified that the liability to HMRC is the total of the tax taken from employees plus all NI contributions, whether employer or employee*
Feedback for incorrect answer: *incorrect, the liability to HMRC is the total of the tax taken from employees plus all NI contributions, whether employer or employee. In your revision, focus on HMRC liability, and what money is due to them.*

Question: *The two entries for the purchase of a computer have been reversed, which of the following will correct the error?*

A: credit bank (twice), debit computer equipment (twice)
B: debit bank (twice), credit computer equipment (twice)

Answer: *credit bank (twice), debit computer equipment (twice)*
Feedback for correct answer: *correct, the original error was to debit the bank account and credit the computer equipment account, whereas what should have happened was: debit computer equipment account, credit bank account. The error is corrected for reversals by performing the correct double-entry transaction twice*
Feedback for incorrect answers: *incorrect, when the error is a reversal of entries, the correct procedure is to perform what should have been twice. The error was: debit bank account, credit computer equipment account. The correct procedure is: debit computer account (twice), credit bank account (twice). Focus your revision on types of error and how each is corrected.*

Question: *Match the recipient with the following payroll-related cost*

Recipients: Employee, HMRC, Pension Company, Employer
Costs: Income tax, NI (employee), Pension contributions, Net salary

Answer: *HMRC – Income tax & NI (employee), Pension company – Pension contributions, Employee – Net salary*
Feedback for correct answer: *Correct, you have identified that the net salary is the final amount payable to the employee, that NI contributions go to HMRC, and that pension contributions go to the pension company (not HMRC)*
Feedback for incorrect answers: *Incorrect, all NI contributions go to HMRC, pension contributions go to a separate pension company, and net salary is the amount payable to the employee. Focus your revision on types of deduction and employer contributions, and practice matching them with their recipient.*

Appendix D – Final Questionnaire

Note: these questions were originally asked and answered online

Question 1: Please indicate your normal working patterns for Tuesdays and Wednesdays by ticking the relevant boxes*, i.e. how many hours a day you would normally work on those days*

No. of hours worked	**Tuesday**	**Wednesday**
No hours worked		
0-2 hours worked		
2-4 hours worked		
4-6 hours worked		
6-8 hours worked		
8+ hours worked		

Question 2: Please indicate by ticking the relevant boxes which methods you used to record and store feedback that you received. You may tick as many boxes as appropriate

Feedback notes method	*Tick if method used*
Make notes on paper	
Print off feedback from VLE	
Take picture on VLE	
Nothing	

Question 3: Please indicate which methods you used in your revision. You may tick as many boxes as appropriate

Revision method	*Tick if method used*
Re-read course materials	
Re-read own notes	
PowerPoints on VLE	
AAT E-learning modules	
AAT Green Light tests	
Course books chapter activities	
VLE activities	
No methods used	

Question 4: Please indicate what effect you felt the pre-written feedback had on your performance in the subsequent retest. *Please tick only one box*

Considerable positive impact	
Some positive impact	
No impact	
Some negative impact	
Considerable negative impact	