Context-dependent memory: Do changes in environmental context cues affect student recall?

Michael Seddon
Liverpool John Moores University

Abstract
This study concerned an action research project undertaken within a mainstream 11-18 secondary school with a high proportion of pupil premium students in the north of England. Pupil premium is a grant given by the government to schools in England to decrease the attainment gap for the most disadvantaged children, whether by income or by family upheaval. The purpose was to investigate whether context-dependent memory impacts student recall during examinations. Students were tested within their standard classroom environment, then moved to a different environment for their second test. The results of this were statistically analysed and compared between genders and school years. The study demonstrated an impact, with students performing statistically worse when tested in an area that is removed from their standard environmental classroom context. Gender was shown to have no impact upon the effects, however, the school year was. Year 7 students were less affected than all other years. The reasons for this are unclear. There were limitations within this study, primarily with ensuring the examination papers were similar enough to act as a control variable. With the range of new topics introduced between the two sets of exams, students had a greater breadth of required knowledge. It was plausible therefore that there were other factors influencing the students’ poorer performance. More research will need to be undertaken to establish that is the change in context that causes lower performance.

Keywords
Context-dependent memory; Memory; Examination; Assessment of Learning; Retrieval.

Introduction and Background
The following investigation was undertaken within a non-academy comprehensive 11-18 secondary school in the north of England with a high proportion of pupil premium (PP) students. Pupil premium is a grant given by the government to schools in England to decrease the attainment gap for the most disadvantaged children, whether by income or by family upheaval. OFSTED (The Office for Standards in Education) had identified an issue in attainment of PP students when compared to their non-PP counterparts.

Context-dependent memory describes the phenomenon in which recall is stronger when a subject is present in the same environment in which the original memory was formed (Hupbach et al, 2008). It has been discussed within psychological literature since the 1950s. Many studies have demonstrated this effect, such as Grant et al (1998) although some have demonstrated more limited information regarding recall (Coveney et al, 2013; Hupbach et al, 2008). It should be noted that the method used by Hupbach was not standard practice in context-dependent studies, making use of interpolated learning which may have impacted on the results.

Context-dependent memory involves episodic memory and retrieval cues. The premise is that memories require a retrieval cue to enable recall (Isarida & Isarida, 2014). The retrieval cue is not limited to the semantic information surrounding the specific target memory, (Tulving & Thomson; 1973; McDermott & Roediger, 2013). It can be any stimuli that are able to be encoded alongside the

Citation
specific target memory. It has also been demonstrated that increasing the overlap between the environmental conditions during encoding and retrieval can significantly aid in retrieval. This is known as the encoding specificity principle (Bramao, Karlsson & Johansson, 2017). The subject of context-dependent memory within education arose from a discussion regarding students undertaking examinations in the hall. If the assumptions of context-dependent memory are correct and applicable, would it be that students are at a disadvantage undertaking their examinations away from the stimuli present when the information was initially learnt (as with a GCSE (General Certificate of Secondary Education) examination)? Would, therefore, examining students in an area with the same environmental stimuli result in less reliable data compared to sitting them away from said stimuli?

Abernethy (1940) explored whether the removal of participants from the formal laboratory used for initial psychological studies would impact their attainment. Abernethy’s study lasted five years, with the initial three years of testing designed to establish a consistent level of difficulty. 181 undergraduate students were used and assigned to groups based on the results of a prior subject knowledge test. It was not, however, specified how that was undertaken. This study involved students doing external reading, which is an aspect beyond the researcher’s control which could have affected the results. The study claimed effective control by ensuring the four conditions of recall were consistent (vision, audition, other senses and nervous system), however as external reading was not controlled by the researcher this is not a valid statement.

Godden and Baddeley (1975) were the primary researchers involved in forming the notion of context-dependent memory. At the time, the evidence base was “far from convincing” (1975, p.325). A group of 18 divers sourced from a university diving club were shown to have greater recall of information when questioned within the area that they were initially taught i.e. on land or underwater. This was a much more significant environmental change than one faced by secondary school students, however other studies demonstrated the same effect in a scholastic setting. The study attempted to ensure all subjects began each test in the same conditions and controlled for aspects such as breathing causing muffling on audio underwater. The amount of diving experience the participants had is not stated, and the amount of familiarity may have influenced the results. It may be that if the experiment was repeated with people who had no experience in diving, or a much larger sample size the clear correlation may not have been shown.

Fernandez and Glenberg (1985) and Eich (1985) demonstrated a lack of a link between incidental context and strength of recall. Fernandez and Glenberg made use of an incidental learning procedure, in which they did not inform the subjects they were specifically testing their recall until immediately before the test. This allowed them to minimise the subjects chance to revise the material, ensuring all subjects recall was from the same point. Fernandez and Glenberg did not conclude that context-dependent memory was incorrect because of their study, simply that standard experimental methods were not able to reliably produce the effects. Further issues with the research base were discussed by Smith (2014), primarily that the mean effect size was d=0.28, this is considered a small effect, suggesting that context-dependent memory may be a correct concept but that its impact is not as large as stated.

Saufley, Otaka and Bavaresco (1985) researched this idea further, looking to incorporate statistical analysis within their research. Several studies had demonstrated the idea of context influencing exam performance before this point, such as Abernethy (1940) and work by Jensen, Harris and Anderson (1971). Jensen, Harris and Anderson concluded that there was a reduction in recall ability of students from grades 2-12 (English Years 3 to 13) when tested away from their initial learning environment. This research is flawed in the same manner as the Abernethy work, in the lack of effective use of statistical analysis. Saufley, Otaka and Bavaresco (1985), akin to Abernethy, focused upon university level students. However, unlike the prior study, this research was undertaken using a wider range of courses and significantly larger sample sizes for each experiment (they varied, from 175 during one experiment to 822 in another) were chosen. Although the sample size was large, the number of
students from each course that were tested within a different context was a very small proportion. Within experiment one, 253 students participated but only 16 students were moved to a separate room for their testing. The researchers suggest that difficulty in arranging alternative rooms was the reason for this. The results showed a lack of significance when tested. However, given that the alternative context students comprised such a small proportion of the total students, these statistics are significantly underpowered. It may have been that, in experiment one, those 16 students were amongst the top performing within the class. They were selected one per page from an alphabetical listing of students, so this was mathematically plausible. A more even distribution of tested and non-tested students would have resulted in a more reliable conclusion. They did identify that there was a difficulty in differentiating questions that would identify context dependence, being of a purely recall nature. At the secondary level, this would be a much less difficult task due to the simplified nature of the examinations (as stated within the paper, there is considerably more rote memory style questioning at that level than the college level tested). One important distinction brought up by Bavaresco (1985) regarded the difference in rooms, stressing that during a student’s time within education they will naturally be exposed to a variety of rooms. It may be that there is simply not a distinct enough change in the environment within an educational setting to generate the difference in recall, as opposed to the starkly different environment used by Godden and Baddeley (1975).

Later research demonstrated that increasing the number of environmental stimuli positively impacts recall (McDermott & Roediger, 2013), so students undertaking external reading and therefore exposing themselves to a greater number of stimuli would have greater recall than students who do not undertake external reading. Smith (2001) demonstrated that the area studied had an impact upon free recall and that multiple forms of encoding would aid in the recall of target information as opposed to single contexts. Smith described this as a reinstatement effect. This was undertaken using a meta-analysis of 75 studies ranging from 1935 to 1997. Abernethy (1940) showed a marked difference in test scores for certain students. Students deemed to be poor and emotionally unstable were affected by the changing of rooms, instructors and seating arrangements (this would likely link in with the modern definition of pupil premium), those deemed high in scholastic ability showed little to no change across conditions. Potentially those students had high scholastic ability as they were undertaking significant amounts of external revision and therefore increasing their number of available stimuli. No statistics were undertaken upon this data as it was undertaken before statistical analysis became a standard aspect of scientific reporting, rendering conclusions less accurate. The largest decrease in attainment demonstrated within this study was when both room and instructor changed, as would occur within a standard GCSE examination.

My study was designed with the aim of creating a more robust method for assessment, using the suggestions presented in the literature, to enable more effective support and intervention for students, and work towards closing the gap between PP and non-PP students. The reformed GCSE specifications in Science consist solely of examinations with no coursework, therefore the need for effective examination protocol is crucial. The hypothesis here is that moving students away from their standard classroom context will result in poorer performance

**Research Methodology**

For this investigation, all assessment points will be referred to as AP1 (the November assessment point), AP2 (the March assessment point) and AP3 (the June assessment point). This study made use of quantitative data to allow for statistical analysis (Madrigal and McClain, 2012), as this had been identified within the literature review as a weak point in many studies, such as Abernethy (1940). As stated by Albers (2017) and McGrath (2013), quantitative data allows for clear analysis of underlying trends and patterns, which is required to confirm the hypothesis given for this investigation. The data came from a non-academy 11-18 secondary school in the North of England with a high proportion (>60%) of pupil premium students. Both KS4 classes (within this school context years 10 and 11) and KS3 classes (years 7, 8 and 9) were used. KS4 had a variation in examination area pre-determined within their academic assessment points. Two sets of data were available for year 11.
students, with AP1 being undertaken in the sports hall, or isolated non-subject classrooms for SEN students (as per standard GCSE regulations) and AP2 being undertaken within the student’s standard Science classroom. It was decided that GCSE results would not be included in this study (given the minimal time that would have been available for data analysis, and other mitigating factors involved in the official GCSE examinations), so data point AP3 for Year 11 was not included. Year 11 data was analysed and discussed separately.

Years 7-10 undertook AP1 and AP2 within their standard Science classroom. Year 10 undertook AP3 within the sports hall. They also undertook three separate papers at 70 marks each, as opposed to their single 50 mark papers for prior assessments. This was acknowledged when analysing the data. Years 7-9 were internally rotated by the researcher for AP3 and students undertook this examination within a different environmental context. It was decided that classrooms would be limited to Science classrooms in order to place no strain on other departments.

The length of the examinations for Years 7-9 was kept consistent with the AP2 at 60 minutes and 50 marks. The examinations were undertaken with the standard classroom teacher, to remove behaviour management as a mitigating factor. Classrooms were issued with a uniform seating arrangement to act as a control variable as well as to minimise potential cheating. Students were given warning of the movement at least a day in advance to aid in efficiency on the day and to minimise the effects of disruption on recall (Strand, 1970)

Data Collection
It was decided that the data collected would be percentage scores and level gained. Originally raw scores were to be collected; however, later assessments used a reduced number of marks from AP1 so it was determined that this would result in less valid conclusions. Students that missed either AP2 or AP3 were removed from the analysis. Once the data was collected, it was decided to use only the percentage gained, due to the altering of grade boundaries between assessment points. This form of data gathering allowed for a larger sample size than previous studies (n=620). Making use of the assigned assessment point allowed for the use of a personalised assessment for each year group. However, making use of students outside the researcher’s classroom did prevent certain potentially confounding factors being controlled (prior teaching styles, in-class revision) and forced reliance upon departmental teachers proceeding with the study correctly.

Ethical issues were minimal within this report. For this study two gatekeepers were involved (Cohen et al, 2011), the Head of Department and the SENCO (Special Educational Needs Coordinator). Four classes were excluded from the movement (n=72) after these consultations, due to concerns raised regarding students’ special educational needs. All data were anonymised via numerical codes, and analysed within school premises until fully anonymised. Marked scores were triangulated within the department, to ensure teachers were marking accurately and enabling valid results.

Data Analysis
Percentage scores were compiled within an Excel spreadsheet, coded by year and gender, and filtered using conditional formatting to display increases and decreases based on prior scores via a colour change. These scores were then coded with a 1 for an improved percentage, a 2 for a decreased percentage and a 3 for no change. Statistical analysis was then performed to interpret the data. Both academic year and gender were gathered and analysed to identify any link between these factors and context-dependent memory.

Presentation of findings
Of the 620 total students (Years 7-10) 257 (41.45%) achieved a higher percentage from AP2 to AP3 when the environmental context was changed, and 353 (56.94%) students achieved a lower percentage. 10 (1.61%) students achieved the same percentage (Figure 1). Comparatively, between AP1 and AP2 when environmental context remained the same, 309 (49.84%), students achieved
higher and 311 (50.16%) achieved lower, with none achieving the same percentage. The mean score at AP2 was 37.9%, and the mean score at AP3 was 34.43%.

![Figure 1](image.png)

**Figure 1.** Students achieving a higher or lower overall percentage between AP1 and AP2 (Same Context), and then AP2 and AP3 (Different Context).

A paired sample T-test was used to analyse for differences between the means of the AP2 and AP3 scores (Figure 2). The results of the paired sample T-test were $t_{614}=6.824$, $p<0.01$. Therefore, indicating statistical significance between the two data sets. A positive correlation between the data sets was demonstrated ($r=0.719$, $p<0.001$). A Cohens d effect size calculation was undertaken and gave a result of $d=0.204$.

When the students that did not participate in the room change were removed from the total ($n=48$), 220 (40.15%) students achieved a higher percentage and 323 (58.94%) students achieved lower. 8 (1.46%) achieved the same percentage. The students that did not move (66) had 37 (56.06%) students gain a higher percentage, 30 (45.45%) gaining lower and 2 (3.03%) performed the same.

Year 11 undertook AP1 in a different environmental context (the sports hall) and AP2 was undertaken within their standard classroom context. Of the 174 total students, 92 (52.87%) improved upon their AP1 scores and 82 (47.13%) performed lower.
Year Variation

The data were also analysed comparatively between year groups, to establish whether the effects of context-dependent memory differed between years (Figure 3). Chi-squared tests were used to test for significant difference between both genders and school years.

The results of the chi-squared analysis showed a significant difference between the years with p=<0.001. Year 7 students were an exception to the trend, with the majority attaining a higher grade after the context swap.

![Figure 3. Percentage change between AP2 and AP3.](image-url)
Gender Variation
The researcher’s school was a mixed gender school, but gender was still analysed to examine whether e.g. split classes would be of a benefit to students (Figure 4). The results of the chi-squared analysis for gender demonstrated no significant difference overall when the environmental context was changed with p=0.420. There was variation in proportions of male and female students in both years and classes, but there was still no trend demonstrated.

<table>
<thead>
<tr>
<th>Year 7</th>
<th>Year 7</th>
<th>Year 8</th>
<th>Year 8</th>
<th>Year 9</th>
<th>Year 9</th>
<th>Year 10</th>
<th>Year 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Higher %</td>
<td>62</td>
<td>56</td>
<td>29</td>
<td>26</td>
<td>33</td>
<td>26</td>
<td>17</td>
</tr>
<tr>
<td>Same %</td>
<td>0</td>
<td>20</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Lower %</td>
<td>30</td>
<td>81</td>
<td>42</td>
<td>39</td>
<td>71</td>
<td>27</td>
<td>69</td>
</tr>
</tbody>
</table>

Figure 4. Gender split percentage change between AP2 and AP3

Discussion and Interpretation
The results of this study showed that students undertaking examinations away from the standard classroom context performed significantly poorer than students tested in their standard classroom context. Excluding data of students who did not change rooms during the study further increased this effect, giving more credibility to the hypothesis. The students that did not change room had a ratio much closer to the overall ratio in AP1 to AP2, potentially due to them maintaining their environmental stimuli. This supports the experiment of Godden and Baddeley (1975) as well as with the wider idea of context-dependent memory. There was demonstrated to be no statistical difference between genders, but different years were significantly affected by the change in context.

The reasoning for Year 7 students not fitting the trend was not clear. It may be that the students took longer to become accustomed to Secondary examinations, so their AP1 scores were a lower benchmark than the other years. Aslan et al (2010) examined how context change affected memory in both children and adults, with their research suggesting that younger children were not influenced by the change in the environmental context in the same way as adults. The Aslan et al study demonstrated that by the fourth grade (Year 5) the effects of context were like that of an adult. It is possible, given the small age gap between Year 5 and Year 7 that some of the students had not developed to a point where changes in context affected them. Given the effects demonstrated on older students however, introducing the movement of classrooms for assessment to the younger students would be a suitable suggestion to allow for familiarisation.

In line with the issues discussed by Smith (2014) within the literature review, Cohen’s test on the data gave an effect size of d=0.204, suggesting a small effect size. It may be that extraneous factors may have a greater effect on students’ memory and attainment than the altering of environmental context.
Much of the studies referenced regarding context-dependent memory were undertaken within laboratory settings. As discussed by Saufley, Otaka and Bavaresco (1985), environmental cues encompass every part of the environment. This can include lighting, wall decorations, even the specific place someone is sat. Within the school context, Science classrooms are required to have several consistent features (e.g. displays) and the layout is identical amongst all but two of the classrooms. It is plausible therefore that the contextual difference was simply not large enough with the Years 7-9 to produce the level of effect seen by Year 10. As stated by Saufley, Otake and Bavaresco (1985), students are potentially accustomed to movement and change of classroom, therefore, the effect of this on retrieval would not be clear. Students move classroom up to 6 times per day in the researcher’s school, and it may be that if this investigation was repeated in a school in which the students did not move (for example a primary school) then the results would be more pronounced. It is also possible that if students were moved to other subject classrooms, as was originally considered, then the effect would also be more evident. Conversely, as mentioned by Coveney et al (2013) and initially proposed by Strand (1970), it may be that the difference comes from the disruption of the room changing. However, the results in the Strand (1970) paper were attained from having students being removed between learning lists as opposed to between learning and recall. Disruption was kept as minimum as possible during the researcher’s study, with students being informed ahead of time as opposed to being moved after arriving at the standard classroom.

The Year 10 students undertook mock GCSE examinations, and Year 11 were preparing for their GCSEs and were theoretically engaging in more extra-curricular revision than Years 7-9. This would likely take place in multiple areas, increasing the number of environmental stimuli and thereby the number of triggers they have to aid recall. This should result in the students having a greater recall and a greater performance. The data clearly demonstrate that this is not the case. However, compared to the assessment that Year 10 undertook for both AP1 and AP2, at AP3 the students undertook three separate 70-mark papers, totalling 210. It may be the larger scope and breadth of each paper resulted in the students performing worse than they would have had they been given a ‘standard’ 50-mark paper. A suggestion would be to ensure that students are always sitting longer, separate GCSE style papers if possible, allowing them to become accustomed to the style and requirements earlier on. Year 11’s reason for not fitting the pattern is unclear.

As demonstrated by Grant et al (1998), there was some degree of difference for participants tested in a noisier or quieter environment to which they initially learnt the information. Examination halls are quieter than classrooms, and many students will study externally to music. Whilst papers were undertaken in examination conditions, volume was not controlled within this study and may have had an impact upon the results (i.e. people walking in the corridor).

Bramao, Karlsson and Johansson (2017) showed that mental reconstruction of the original encoding context significantly improved subjects recall. Masicampo and Sahakyan (2014) demonstrated a similar process, with the imagining of another context during the encoding stage being able to counteract the effect of context-dependent forgetting. With the decrease in test scores demonstrated for many students, it may be that investigating the effects of this mental reconstruction or imagination would be beneficial.

One aspect that may have influenced context-dependent memory was mood, as suggested by Xie and Zhang (2017). They worked from a hypothesis from research reviewed by Surprenant and Neath (2009) that successful retrieval is more likely when the emotional states are matched during both encoding and the attempted retrieval. Their research suggested this was accurate, with participants performing better when the emotional state was consistent between the encoding and testing phases. However, it should be stated that only 40 participants were used in the study. This was also not undertaken in a written or oral context, akin to a GCSE examination. Instead, participants were recalling specific shades of colour that an object was in an originally learnt picture.
Implications for Future Research and Practice
Moving forward, there are several areas that practitioners can work on in order to enable students to start to overcome the issues discussed within this article. Ensuring students are exposed to the greatest number of external stimuli possible would be of great benefit. Practitioners should encourage students to revise in multiple areas, and if possible, change classrooms on occasion for e.g. intervention sessions or after school revision. As McDermott and Roediger (2013) suggested, this should positively increase their recall, perhaps counteracting the impact of the room change. From the research of Bramao, Karlsson and Johansson (2017 and Masica and Sahakyan (2014), having students practice mental reconstruction through memory techniques such as the method of loci, or picturing exam halls during initial encoding could potentially aid students in counteracting the impact of the change of context. Students would likely find the emotional state of their exam to be more negative than the initial learning or revision (Zie and Zhang, 2017). Working on ensuring students feel adequately prepared (practising papers, undertaking more room changes) may help to reduce the negative feelings and result in a more consistent mood across the two scenarios. The department will be modifying its examination procedure based on the results of this study, as well as further considering aspects such as classroom swaps for lessons or revision sessions.

Conclusion
Students in a school context such as the researcher’s (high PP) historically perform worse than more affluent areas. It is vital for those students that strategies are put in place to enable them to achieve as well as their non-PP counterparts. Understanding the impact of factors such as context-dependent memory can allow for more accurate data collection, and therefore a more effective academic intervention to help close the gap, as well as improve scores for all students.

This study was intended to establish whether a change in the environmental context between students learning and testing would affect their recall in line with the theory of context-dependent memory. Students were tested in their standard classroom context, and then tested in a different context and the data was compiled and analysed. Within the scope of this report, the results confirmed the theory of context-dependent memory, and students were demonstrated to have performed significantly poorer when placed in a different classroom environment. Year 7 students were the exception to this rule, the reasoning for this is unclear, but suggestions were discussed. The limitations within this methodology have been discussed, such as the implementation of new topics and the difficulty in keeping tests similar enough to draw valid conclusions. Further areas for research and suggestions were given, such as whether pupil premium status impacts the effects of context-dependent memory.

References
Abernethy, E.M. (1940) The Effect of Changed Environmental Conditions upon the results of college examinations. The Journal of Psychology. 10(2) 293-301


McGrath, A. (2013) *Reader’s Response Describing and Analyzing Quantitative Data.* International Journal for the Scholarship of Teaching and Learning. 7(1) 1-4


