Students pay attention!

Combating the vigilance decrement to improve learning during lectures

MARK S. YOUNG, STEPHANIE ROBINSON AND PHIL ALBERTS Brunel University, UK

ABSTRACT Maintaining student concentration in lectures has long been a challenge for lecturers. Pedagogical research consistently finds a drop in attention between 10 and 30 minutes into the lecture, which has been associated with the passive nature of the standard format, and has consequences for learning approaches and outcomes. A similar phenomenon has been observed in ergonomics for some time, known as the vigilance decrement. In this article, we present an exploratory effort to detect the vigilance decrement in four different lecture formats, by adopting an ergonomics measurement tool which has been related to vigilance, and relating the findings to students’ assessment results. It was found that standard lecture formats do induce a vigilance decrement, and this can adversely affect learning of the material. Conversely, vigilance degradation is avoided when presentation is varied, though this is not necessarily associated with interactive participation techniques. Implications for lecturing styles, learning approaches and pedagogical research methods are discussed.

KEYWORDS: active learning, attention, deep and surface learning, ergonomics, lectures, pedagogy, vigilance

Introduction

University teaching popularly and historically revolves around the traditional ‘chalk-and-talk’ lecture method. Indeed, the efficiency of lectures means that they continue to be the primary method of instruction, particularly for large class sizes (Heward, 2003; Lammers and Murphy, 2002). Nevertheless, doubts remain not just about the efficacy of lectures in achieving learning outcomes, but also in terms of their popularity amongst students (Huxham, 2005). One of the problems with traditional lectures is that students are in a passive mode of learning (Heward, 2003), which
adversely affects their attention and their ability to retain information (Windschitl, 1999). Such passivity has two potential consequences for the student; it does not facilitate deep learning (e.g., Gibbs and Habeshaw, 1989; Lammers and Murphy, 2002), and it can cause decrements in student concentration. Thus we have seen a recent upsurge in active learning techniques which, in the present article, we consider to be those that encourage students to ‘actively participate in the learning experience rather than sit as passive listeners’ (Lammers and Murphy, 2002: 55).

As passive listeners or monitors, people generally find it difficult to maintain a constant level of attention over extended periods of time. The nature of the traditional lecture means students are particularly liable to concentration problems (Bligh, 2000; Gibbs and Habeshaw, 1989). Various studies have found that attention degrades after between 10 and 30 minutes on task (Frederick, 1986; Horgan, 2003; Stuart and Rutherford, 1978). A similar phenomenon has been established in the field of ergonomics – the vigilance decrement.

Vigilance tasks involve monitoring for low frequency signals from a background of noise (Grier et al., 2003), and it has been known for some time that it is virtually impossible to maintain high performance on such tasks for any length of time (Mackworth, 1948). The traditional radar operator’s task is a classic example, but in recent times vigilance issues are more likely to arise with human supervisory control of automation, a task at which humans are particularly poor (Parasuraman, 1987). As with lectures, these monitoring tasks represent a shift from active to passive information processing (Endsley and Kiris, 1995), and performance decrements similarly occur within 20–30 minutes (Singleton, 1989; Warm et al., 1996). Contrary to popular opinion, though, recent research suggests that the mental demands of sustained vigilance are quite substantial (Warm et al., 1996).

Various authors advocate mixing up the level of stimulation during lectures in order to offset the vigilance decrement. The demands should be changed every 10–15 minutes (Horgan, 2003; Wankat and Oreovicz, 2003), and the source of such variety could be a simple rest, a change in presentation medium, or setting the students a short task (Bligh, 2000; Frederick, 1986; Race and Brown, 1998). As with vigilance studies in human supervisory control, these interruptions only temporarily restore attention levels, and afterwards concentration will decline even more steeply (Bligh, 2000; Gibbs and Habeshaw, 1989). Taking this a step further, then, the active learning philosophy uses constructive and task-related activities to break up the lecture. Typically, this will involve a small group discussion or ‘buzz group’ (Gibbs and Habeshaw, 1989;
Windschitl, 1999), some structured debate (Kumar, 2003), or even novel technological forms of interaction (Beekes, 2006; Rosie, 2000). Part of the motivation behind active learning techniques is that they are assumed to facilitate deep learning approaches in students (Horgan, 2003).

Clearly, then, there is a good deal of overlap between applied vigilance research in ergonomics, and the problem of student concentration in the pedagogical domain. The respective solutions are also common – reset the vigilance decrement or restore some level of active processing into the human’s task. Thus there are direct parallels between active learning in the classroom and active operations in the control room. Nonetheless, there remains a question as to whether active learning techniques have advantages over and above simply resetting the vigilance decrement. In other words, for deep learning, do we need to ensure that our students are maintaining attention and actively participating, or is it sufficient that our teaching methods are designed only to keep attention up (which can be achieved simply through strategies such as short breaks)? Nobody has yet directly tested the relative effectiveness of just breaking the lecture up against interspersing it with activities.

In the current article, we present an exploratory study to illustrate the efficacy of a structured programme of research into these issues, and propose a methodology for such an investigation. We adopt a vigilance measurement technique from ergonomics to monitor students’ attention levels during four different lecture formats, and compare the results with their performance on an end-of-term assessment. By assessing vigilance levels and manipulating the level of active learning, it should be possible to distinguish the influence of these factors in quality of learning outcomes.

**Method**

Like most pedagogical research, the present study relied upon a naturalistic paradigm and this, coupled with the ethical implications of manipulating teaching methods for research, necessitated a quasi-empirical design to assess whether student learning in lectures is subject to a classic vigilance decrement. Essentially, this is an ethnographic approach combining quantitative methodologies from the field of ergonomics with field observations in a pedagogical environment. While it would not be reliable to generalize the findings completely from such a study, the data will be analysed for trends and indicators that merit further research.

**Design**

The study focused on a consecutive series of five one-hour lectures during a final-year module on an undergraduate degree programme. Four different
lecture formats were used during the study. In the first two lectures (L1 and L2), the traditional ‘chalk-and-talk’ method was adopted, with the lecturer addressing the class via PowerPoint slides with no opportunity for rest or interaction. Lecture 3 (L3) introduced a guest lecturer, whose material remained integral to the module but was still presented in a standard lecture format. In lecture 4 (L4), the session was paused about halfway through the lecture for a short ‘buzz group’ discussion session. Finally, lecture 5 (L5) departed from the standard format by using a range of case studies, some of which were illustrated using video media.

The rationale behind these lecture formats is to tease out the influence of active learning as opposed to simply varying the stimulation for the students. In the standard lectures (L1 and L2), it is expected that vigilance decrements will occur, and that deep learning will be inhibited. Both the guest lecturer (L3) and the case studies lecture (L5), in introducing a level of novelty, may partly attenuate the vigilance decrement, though it is not anticipated that this would strongly affect the learning approach. Gibbs and Habeshaw (1989) suggest that guest lecturers can increase the effectiveness of lectures, while several authors support the use of technology (i.e. format L5) to vary stimulation in the lecture (Frederick, 1986; Gibbs and Habeshaw, 1989; Lammers and Murphy, 2002). Using buzz groups in lecture 4 should increase active (and therefore deep) learning as well as offsetting the vigilance decrement (Gibbs and Habeshaw, 1989; Horgan, 2003).

**Measurement**

In ergonomics, a widespread tool for measuring mental workload is the NASA Task Load Index (TLX; Hart and Staveland, 1988). The TLX is a multidimensional subjective rating technique with six subscales (Mental Demand – MD, Physical Demand – PD, Temporal Demand – TD, Performance – PE, Effort – EF, and Frustration – FR). Participants rate each subscale by making a mark on a visual analogue scale, which is accompanied by a definition of the subscale. For instance, Mental Demand is defined thus: ‘How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?’ Each scale is anchored from low to high, except Performance, which uses good to poor (as a rating of one’s own performance) – the logic being that a low workload task would elicit all responses at the same end of the visual analogue scales.

The subscales are then scored from 0 to 100, and the mean of these scores constitutes overall workload for the task. Previous work has found
that vigilance tasks evoke moderately high ratings on the TLX (Grier et al., 2003). Moreover, Warm et al. (1996) found that there is a signature pattern of responding on the TLX which is indicative of a vigilance task. This ‘vigilance footprint’ is characterized by the primary contributions of the Mental Demand and Frustration subscales to the overall workload score (see Figure 1 for an example, drawn from a study on drivers using vehicle automation; Young, 2000).

The TLX was administered at two points during each lecture – once around halfway through (time 1), and once again at the end (time 2). Note that, for L4, the TLX at time 1 was administered shortly after the buzz group discussion. Typically, the TLX would be anchored to some baseline activity to ensure consistency across responses. However, such anchoring is unnecessary when making relative judgements on the repeated-measures design used here.

The five lectures under study here formed part of an eight-lecture section of the module, after which an end-of-term summative assessment was conducted. The assessment, essentially a three-hour exam, consisted
of two parts: in the first hour, students answered a series of multiple-choice questions (MCQ) with a four-option forced-choice response. There were four MCQs for each lecture topic. The remainder of the exam was an elective essay component, with students selecting two from eight questions (i.e. one essay per lecture topic). The overall assessment contributed 40% of their marks for the module.

**Participants and procedure**

There were 49 students (38 male) in the class, all of whom sat the end-of-term exam. However, since the TLX was administered as a voluntary exercise, and class attendance varied from week to week, not all of these participants completed every questionnaire, and thus sampling was very much on an opportunity basis. Consequently the data points for the TLX varied from $n = 24$ to $n = 37$ and, while all students completed the multiple-choice questions, the optional essay questions ranged from $n = 1$ to $n = 28$. Given these constraints on the within-subjects design, in the statistical analyses which follow, missing cases are treated in a listwise fashion, hence the varying degrees of freedom in each test. Average age of the whole class was 21.9 (SD = 1.52).

In terms of administering the TLX, response sheets including definitions of the scales were handed out in hard copy at the start of the lecture and completed on instruction from the lecturer, at an appropriate point as close as possible to the midpoint of the lecture, and again at the close of the lecture. The instruction simply requested students to complete the TLX based on their perception of the previous section of the lecture (i.e. at the close of the lecture, students only rated their perceptions since the last TLX), with reference to the definitions of the scales. The instructions were repeated consistently on each occasion, and students were given around two minutes to complete the response sheets.

**Results**

Parametric statistical tests were used throughout the data analysis, using paired t-tests, analyses of variance (ANOVA), correlations and regressions as appropriate to test the hypotheses. Dependent variables for the TLX were primarily focused on the subscales to identify the vigilance footprint, but overall workload (i.e. the arithmetic mean of the subscales) was also used to correlate with the assessment results. Dependent variables for the assessment grades were the percentage points for the multiple-choice and essay tests.
Mental workload
A paired-samples t-test comparing the overall workload scores from time 1 to time 2 across all lectures revealed a significant increase during the course of the lecture, from 44.7 at the midpoint to 49.1 by the end ($t(43) = -2.449; p < 0.05$).

Breaking this analysis down between lecture formats, two separate analyses of variance (ANOVAs) were conducted, one each for time 1 and time 2. At time 1, overall workload significantly differed across the lectures ($F(4,52) = 3.295; p < 0.05$), with L1 and L3 constituting the source of the difference. As can be seen in Figure 2, workload is considerably lower for L3, while L1 imposes a higher demand. The ANOVA at time 2 was non-significant ($F(4,40) = 1.521; p = n.s.$).

Assessment
An ANOVA across lecture formats for the MCQ component returned a significant result ($F(4,192) = 12.02; p < 0.001$), with both of the standard lectures standing out in different ways. While results on the L1 topic were significantly lower than the rest, L2 scores were significantly higher (see Figure 3).

![Figure 2](image-url) Overall workload scores across each lecture at time 1 (midpoint)
As students were given a choice of essay questions, there were not enough data to compare these results statistically. Nonetheless, a visual inspection (Figure 4) shows a clear dip in results for the material delivered by the guest lecturer (L3). Perhaps more telling is the popularity of each topic – with only one student responding to the essay for L1, and three students attempting the L3 essay.

A series of correlation analyses were also carried out between the TLX data and the assessment results, to determine whether there was any association between perceptions of workload in the lectures and quality of learning outcomes. A few relatively weak but significant correlations are worthy of note. Firstly, there is some evidence of an association between workload measured at the midpoint of L1 and the MCQ results for that lecture’s topic ($r = 0.32; p < 0.1$). A similar correlation was found in L5 ($r = 0.37; p < 0.1$). For the essay questions, small sample sizes generally prevented statistical analysis, but a couple of tests were possible and yielded significant results. Workload at the end of L5 was negatively related to scores on the essay questions for that lecture ($r = -0.696; p < 0.05$). Apart
from the workload data, it was also found that the MCQ scores and essay marks were correlated for the L2 topic ($r = 0.477; p < 0.05$).

**Vigilance decrement**

Finally, a regression analysis on the TLX data was used to determine the presence or absence of the vigilance footprint during each lecture. Although the footprint was not as dramatic as has been observed elsewhere (Warm et al., 1996; Young, 2000; see Figure 1), it was apparent in both of the standard lectures. The standardized beta coefficients (Table 1) demonstrate that particularly at time 1 (halfway through the lecture), mental demand and frustration figure highly in the contribution to overall workload. In contrast, this footprint is distinctly absent in the remaining analyses.

The overall workload scores of between 40 and 50 are not incomparable with previous data suggesting a vigilance task imposes high demands (Grier et al., 2003) found scores around 50–60 in their study). Thus, as has been found in ergonomics (Grier et al., 2003; Matthews et al., 1993; Smit et al., 2004), concentration problems in the lecture theatre may be more due to depletion of attentional resources over time than due to boredom or underload. Furthermore, the higher workload score of L1 and its closer

*Figure 4* Mean essay scores for each lecture format (out of a maximum 35 marks)
bearing to the vigilance footprint in the subscales are suggestive of the fact that this lecture was most representative of the vigilance decrement.

**Discussion**

One of the key results to emerge from this study was that a standard ‘chalk-and-talk’ lecture format does appear to cause the classic vigilance ‘footprint’ as identified by Warm et al. (1996) on the NASA-TLX, while the novel lecture formats did not evoke the same pattern of responses. This was particularly evident at the halfway point in the lecture (after around 30 minutes), which reflects previous research both in vigilance and student concentration. That the decrement is less apparent at the end of the lecture is also consistent with the ‘mid-lecture dip’ (as reported by Bligh, 2000).

In terms of learning outcomes, there are several intriguing comparisons to be made. Firstly, despite both standard lectures showing signs of a vigilance decrement, the exam results for L1 and L2 swing from one extreme to another. MCQ scores on L1 are significantly lower than the other lectures, while those for L2 stand out above the rest. Meanwhile, although essay scores on both are similar, the numbers of responses to each question are revealing – only one for L1, while L2 is the most popular topic, with 28 students attempting it. The natural conclusion to draw from these data is that the material taught in L2 was more popular than that in L1, a conclusion supported by informal feedback from the students.

So, although the standard lecture format can induce vigilance problems, if the material itself is stimulating enough then student perceptions and approaches to learning could still be influenced. While we did not

<table>
<thead>
<tr>
<th></th>
<th>MD</th>
<th>PD</th>
<th>TD</th>
<th>PE</th>
<th>EF</th>
<th>FR</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time 1</td>
<td>0.275</td>
<td>0.220</td>
<td>0.269</td>
<td>0.219</td>
<td>0.276</td>
<td>0.325</td>
</tr>
<tr>
<td>time 2</td>
<td>0.236</td>
<td>0.195</td>
<td>0.244</td>
<td>0.213</td>
<td>0.239</td>
<td>0.313</td>
</tr>
<tr>
<td>L2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time 1</td>
<td>0.312</td>
<td>0.335</td>
<td>0.284</td>
<td>0.256</td>
<td>0.276</td>
<td>0.345</td>
</tr>
<tr>
<td>time 2</td>
<td>0.207</td>
<td>0.307</td>
<td>0.261</td>
<td>0.269</td>
<td>0.278</td>
<td>0.310</td>
</tr>
<tr>
<td>L3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time 1</td>
<td>0.253</td>
<td>0.243</td>
<td>0.283</td>
<td>0.264</td>
<td>0.293</td>
<td>0.285</td>
</tr>
<tr>
<td>time 2</td>
<td>0.240</td>
<td>0.283</td>
<td>0.226</td>
<td>0.210</td>
<td>0.258</td>
<td>0.280</td>
</tr>
<tr>
<td>L4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time 1</td>
<td>0.239</td>
<td>0.327</td>
<td>0.225</td>
<td>0.316</td>
<td>0.308</td>
<td>0.263</td>
</tr>
<tr>
<td>time 2</td>
<td>0.197</td>
<td>0.262</td>
<td>0.300</td>
<td>0.253</td>
<td>0.224</td>
<td>0.253</td>
</tr>
<tr>
<td>L5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time 1</td>
<td>0.238</td>
<td>0.254</td>
<td>0.248</td>
<td>0.201</td>
<td>0.173</td>
<td>0.250</td>
</tr>
<tr>
<td>time 2</td>
<td>0.271</td>
<td>0.318</td>
<td>0.277</td>
<td>0.274</td>
<td>0.234</td>
<td>0.280</td>
</tr>
</tbody>
</table>
record any data on deep or surface approaches to learning, we can make some inferences from students’ strategies on the assessment, and speculate on how the lecture format and material might affect these learning approaches. Our view is that, while manipulating the syllabus material might garner student interest, it is unlikely that such a strategy will ultimately foster deep learning. This conclusion is based on the observation that, although L2 was apparently more popular with students, it only led to benefits on the MCQ component of the assessment, while the essay grades remained at an average level. It is reasonable to assume that such differences in assessment results reflect differences in approaches to learning (see, for instance, the position of the UK Higher Education Academy’s Engineering Subject Centre at www.engsc.ac.uk/er/theory/learning.asp). Several authors suggest indirect evidence to link MCQ tests and essay assessments with surface and deep learning respectively (Clark, 2004; Horgan, 2003; Huxham, 2005; Scouller, 1998), although this is not a universally accepted view (Imrie, 1995; Gijbels et al., 2005; Segers et al., 2006). Nevertheless, given our earlier supposition that L1 posed the bigger vigilance problem, we could suggest that the vigilance decrement severely affects learning across the board, judging by the lower assessment scores for the L1 topic.

Thus we turn to more novel forms of interaction in the lecture, in an effort to both break the vigilance decrement and potentially facilitate deep learning. The three alternative formats used here have been suggested in previous literature as effective in breaking the vigilance decrement – a guest lecturer (as advocated by Gibbs and Habeshaw (1989)), buzz groups (Bligh, 2000; Huxham, 2005), and a more practical and technological approach with video case studies (using alternative media is a method suggested by both Gibbs and Habeshaw (1989) and Lammers and Murphy (2002)). It is notable that only L4, with the buzz groups, includes any real active learning element (at least according to the convention adopted for this article, which includes some interactivity or participation as a prerequisite).

Ironically, L3 given by the guest lecturer resulted in the lowest overall workload results, but also the lowest scores in the essay questions. Again, this was apparently an unpopular topic (only three students attempted the question); moreover, student feedback suggested that the novelty factor of the guest lecturer was both an advantage and a disadvantage – the new presenter’s style (which was itself something of a distraction) partly led students not to concentrate as much as they would normally. Consequently, vigilance was offset and the lecture was rated as easier on the TLX, but students perhaps failed to appreciate the level at which the material was being taught, which then impacted on their assessment results.
Lecture 5, with the video case studies, also used novel presentation styles, and was a reasonably popular topic (17 students attempted the L5 essay question). A dramatic improvement in essay grades was observed, with results for this topic being among the highest of all. MCQ scores, while not the highest recorded, were still significantly better than those for L1. Intriguingly, a very similar set of results was evidenced for the active learning lecture (L4), with both the MCQ and essay scores being comparable with L5. Again, then, we might speculate that improvements in both surface and deep learning could potentially occur just with the right lecture material and a novel presentation style, without necessarily invoking (inter)active learning.

In spite of these findings, the methodological limitations of this study compel us to be cautious in generalizing the findings too far. The ethical and moral constraints of collecting data from a ‘live’ module meant that the confounding influences of subject matter could not be controlled for, nor could we practicably manipulate the lectures to provide a fully balanced experimental design. With this in mind, we would rather offer this article as a foundation for further research in this area – highlighting the need to tease out the relative influences of breaks versus active learning on student vigilance and approaches to learning. Moreover, we have presented the value of adopting a methodology from an analogous discipline in tackling this research question. Future studies would do well to include a wider variety of lecture formats (including breaks with and without activities), with more consistency in topics and delivery, and use measures of deep and surface learning (such as the Study Process Questionnaire: Biggs, 1987). Another interesting aspect to explore would be the nature and structure of the learning material, to see whether the topic matter itself facilitates different learning approaches (as evinced in L1 and L2 of the present study).

Conclusions

The results of this study suggest that student concentration decays in the same way during a passive lecture as does that of a human operator monitoring automated equipment, with serious implications for learning and performance. The recommendations in terms of maintaining attention and concentration are also analogous – instead of interspersing periods of manual control (Parasuraman et al., 1996), short breaks or novel activities may temporarily restore attention to normal levels. Notwithstanding the methodological limitations of the study, and in the light of much previous research in this area (Bligh, 2000; Gibbs and Habeshaw, 1989), we can
conclude that any variation in presentation or media can only help to maintain attention and facilitate deeper learning approaches in all lectures. While buzz groups and other interactive sessions do have clear advantages, there are also some disadvantages such as decreased lecture time, reduced accuracy and lower control over the lecture (Huxham, 2005; Lammers and Murphy, 2002). Alternative ways of maximizing student attention and learning, such as mini-lectures, case studies or videos, might therefore be equally beneficial for learning, yet avoid these disadvantages. While we assumed a ‘strict’ interpretation of active learning, the results of our study implied that true interactivity is not necessarily a criterion for deep learning to occur, suggesting that a broader definition of active learning as ‘a process of engaging with the learning task at both the cognitive and affective level’ (Fry et al., 2003: 432) is probably more appropriate. On this basis, simply breaking the vigilance decrement could be just as effective.

Acknowledgements
The authors are grateful to two anonymous reviewers for their helpful comments and feedback on an earlier draft of this manuscript.

References


Biographical notes

**DR MARK S. YOUNG** is a Senior Lecturer in ergonomics and a member of the Ergonomics Research Group in the School of Engineering and Design at Brunel University.

Address: The School of Engineering and Design at Brunel University, Uxbridge, Middlesex UB8 3PH, UK. [email: m.young@brunel.ac.uk]

**DR STEPHANIE ROBINSON** is a lecturer in learning and teaching on the PG Cert in Learning and Teaching in Higher Education within the Learning and Teaching Development Unit at Brunel University.

Address: The Learning and Teaching Development Unit, Brunel University, Uxbridge, Middlesex UB8 3PH, UK.

**DR PHIL ALBERTS** is Head of e-Learning in the Learning and Teaching Development Unit at Brunel University.

Address: The Learning and Teaching Development Unit, Brunel University, Uxbridge, Middlesex UB8 3PH, UK.