Tubular Daylighting Devices and People

A comparison of the human responses to Tubular Daylighting Devices and fluorescent lighting

By

James Sullivan

A thesis
submitted to Victoria University of Wellington
in fulfilment of the requirements for the degree of
Master of Building Science

Victoria University of Wellington
2013

ABSTRACT

Tubular Daylighting Devices are used to bring daylight into deep-plan spaces, and meet sustainability goals. However, they are expensive, and justification for their use lies in hypothesied benefits they can provide in areas such as well-being and productivity. Yet, there is very little research into the effects of Tubular Daylighting Devices. The broader daylighting literature suggests that benefits to satisfaction, mood, and performance are possible — though research into the benefits of daylight is still not conclusive.

Therefore, a before and after study was carried out in a windowless computer room in the university to compare how the students responded under TDDs versus typical electric lighting. Their cognitive performance, change in mood, average sleepiness, and perceptions of the room and lighting were measured.

TDDs significantly increased ratings of room attractiveness and brightness, and had no more perceived glare than the electric lighting. Ratings of lighting quality were on a par with both typical electric lighting and good modern lighting. They were also just as effective on overcast days as sunny. No effects were found on performance or sleepiness, and mood results were inconclusive.

Overall, it is suggested that TDDs can be considered to be on a par with good modern lighting, and superior to typical existing lighting. Note, however, that it is possible that effects in rooms with windows could differ from those found here. Further research should use longer exposures and larger sample sizes if they wish to find performance effects.

CONTENTS

FIGURES

TABLES

ACKNOWLEDGEMENTS

First, I'd like to thank my supervisor, Michael Donn, for his tireless, and always helpful, advice and guidance.

I would like to gratefully acknowledge the funding provided by HomeTech that made this research possible. I would also like to thank Norman Smith, of the Rocky Mountain Institute, for his comments and his vital role in bringing the university and HomeTech together, and enabling this study. Thanks also to the team at HomeTech for their support: James Hadley, who worked out the design and installation of the Solatubes, and who handily dealt with the complications that that arose; Mike Cullen, the manager, who provided much helpful aid and input.

Thanks too to my parents for their support, particularly my mother for proof-reading the thesis, and my father for teaching me how to do statistical analysis.

I'd also like to thank Andrew Wilks, of the university's facilities management team, who provided the energy monitoring of the room; Jennifer Veitch and Guy Newsham, of the National Research Council Canada, for providing the lighting surveys they used in their studies; Peter Ramutenas, of the IT department in the school of architecture, for his work setting up the light sensors; Stewart Milne, the IT administrator, for helping to set up the systems that allowed the students to get the tests and submit their results over the computer network; My fellow tutors —Anthony Gates, Kate Walker, Jake Osborne, Matt King, Peony Au, Shaan Cory, and Yasmin Merwood — for their assistance in allowing the tests to be run in the tutorials; And my fellow Masters students, for helping to pilot test the tests.

Special thanks should also go to the participants, who provided the data that made this study possible.

1 INTRODUCTION

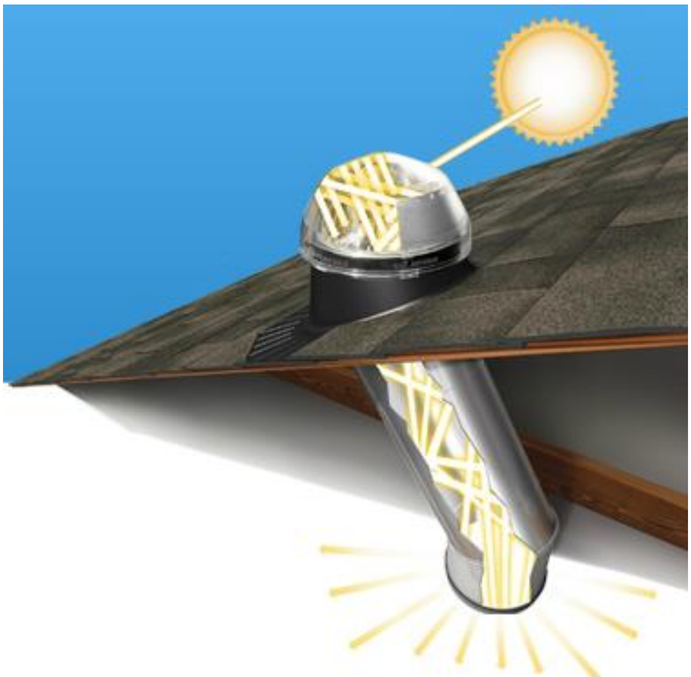


Figure 1-1: The subject of this thesis – The Tubular Daylighting Device (source: Solatube, 2012a)

1.1 Tubular Daylighting Devices — The Problem

> The Tubular Daylighting Device.
>
> A device that can bring daylight into spaces such as deep-plan offices where windows are not an option.
>
> A daylighting system that avoids the problems of glare and excessive solar heat gains.
>
> - Paraphrased from Solatube (2010)

Tubular Daylighting Devices (TDDs) are used in buildings throughout the world to provide natural light, ranging from the residential to the commercial and the educational. In the modern world of sustainability rating schemes like GreenStar and LEED, TDDs are promoted as a way to achieve daylighting and energy efficiency goals.

However, TDDs are expensive. They are marketed on the basis of being able to provide certain benefits, not only in energy savings, but also to mood, health, and performance for the occupants (Solatube, 2012b). Indeed, being able to provide these less tangible benefits to the occupants can be regarded as critical for the value of the TDD: estimates of the energy savings provided by the TDDs installed for this study suggest a payback period of nearly 20 years (see Appendix A). Similar results were found in a European study (Mayhoub & Carter, 2011). If, however, they could provide even small improvements in productivity, then they would become much more valuable, as the cost of the workers (wages etc.) may be as much as 160 times the cost of energy in a building (Mayhoub & Carter, 2011). Similarly, enhancements to student learning in schools could also have great value.

If TDDs *can* provide such benefits, in areas such as well-being and productivity, then they could be a valuable way to be more sustainable, improve the economy, and provide a better environment for people.

So, do Tubular Daylighting Devices provide such benefits?

1.2 There is a paucity of research into their effects

Unfortunately, human response to the use of Tubular Daylighting Devices in buildings is an area with a severe paucity of research. The vast majority of research about TDDs has focused around the technical aspects of their design and implementation. There are only two studies of how people respond to them, both of them field studies of office buildings (Carter & Marwaee, 2009; Marwaee & Carter, 2006). There is also another loosely related field study of two *active* daylight guidance systems (Ejhed, 2001), but it has limited application here.

1.2.1 The available research tells us little

Ejhed found that people preferred daylight, could detect changes in outdoor conditions through the lighting systems, and could identify daylight from its colour (as cited by Marwaee & Carter, 2006). However, as he was evaluating systems other than Tubular Daylighting Devices, his study cannot say how well people respond to TDDs.

Carter and Marwaee (2009) surveyed workers in 22 buildings in the UK that had been equipped with TDDs. They found that:

- People *can* identify the light provided by TDD as daylight. However, they do not consider its quality or quantity to be as good as that provided by windows.
- People are more satisfied with the lighting when there are windows in the room, even if they are not making a discernible difference to the light levels. In rooms without windows, people tend to be dissatisfied with the lighting.
- Around 25% to 33% of users could detect weather and time through the light in the windowless rooms. This suggests that TDDs can provide this outdoor connection to a degree, but are far less effective than windows.
- Perception of the amount and quality of daylight improves when more daylight is provided.

They noted however that the studies were limited by the fact that the TDDs in the buildings generally provided significantly less light than the electric lighting, and hypothesised that improved effects might be found in a "well daylit space" with a higher daylight factor (>2%). They concluded that, while offices with windows and TDDs are better than those without windows, there is still not enough evidence to say whether or not offices with TDDs are better than those without (Carter & Marwaee, 2009).

It should also be noted that their studies focused on people's perceptions of the lighting. Thus, the areas of TDD's potential effect on factors such as performance, mood, and health are still unstudied.

1.3 Aim of this study

This study examines the question of whether or not Tubular Daylighting Devices provide benefits to people that would justify using them over other forms of lighting.

Other forms of lighting are 1) traditional daylighting (i.e. windows) and 2) electric lighting. As one of the key advantages of TDDs is their ability to reach places that traditional daylighting can't, electric lighting can be considered one of their main competitors. Thus, this study focuses on the comparison between TDDs and electric lighting.

To that end, several key questions need to be asked. The first question is:

"What evidence is there that TDDs could provide benefits to people beyond electric lighting, and what benefits could there be?"

This is answered through a review of the daylighting literature in Section 2.

Once it has been established that there is reason to test TDDs to identify possible effects , the next question that needs to be asked is:

"What is the most appropriate way to assess the potential benefits of TDDs?"

This is addressed in Section 3, which details the development of the testing methodology.

Once these issues are addressed, the tests and comparisons between TDDs and electric lighting can be run (Sections 4 and 5), allowing us to answer the main question of this study, specifically:

"Do people respond better to Tubular Daylighting Devices than to electric lighting in offices and educational facilities?"

2 DAYLIGHTING AND PEOPLE: THE CASE FOR TUBULAR DAYLIGHTING DEVICES

The following chapter is a review of the literature on the human response to daylight[ing], for the purpose of determining what we know about the effects of daylight[ing], in order to determine what kind of benefits TDDs may provide, as well as how credible the notion of them being better than electric lighting is.

2.1 The questions

As has been discussed, studies into people's response to TDDs are very limited. Despite this, however, TDDs are marketed on the beneficial effects of their light. This is done by referring to the purported benefits of *daylight* for people. The logic is that TDDs provide daylight, and that daylight is good.

The key questions here for evaluating the case for TDDs in this regard are:

1) Has daylight been shown to be beneficial?
2) If daylight is beneficial, are the mechanisms by which it provides benefits present in TDDs?

If the answers to these questions are positive, then there would be a case for arguing that TDDs are likely to be beneficial to people, despite the lack of specific research into TDDs.

2.2 A note: Daylight vs. Daylight*ing*

It is important to note the difference between daylight — the light provided — and daylight*ing* — the means used to provide the daylight. Most studies comparing daylight and artificial lighting are also comparing the different lighting strategies. Daylight, as provided by a window, is very different to artificial light as provided by ceiling mounted luminaires. This means that if, say, a study found that daylighting was better than artificial lighting then there would be a large number of potential reasons for it. For example, it could be that daylight is inherently better than fluorescent light, or that the view through the window is good, or because windows tend to provide much higher illuminances than electric lighting, or because the light from the windows was coming from the side rather than from above. These are just some of the possible explanations. Most of the studies into daylighting are thus limited in their ability to have their results applied to TDDs as they differ from windows in many ways.

2.3 The claims: Purported benefits of daylight[ing]

Daylighting is generally preferred to artificial lighting and is widely believed to be superior (Boyce et al., 2003). In light of this, many claims are made about beneficial effects of daylight (e.g. Solatube, 2012b). It is claimed that daylight can:

1) Improve productivity in the workplace
2) Improve mood and general well-being
3) Reduce absenteeism
4) Improve quality of work
5) Improve job satisfaction

6) Provide higher sales and greater customer loyalty
7) Help students to learn faster and get higher test scores
8) Make spaces appear more appealing

So, does the evidence support these kinds of claims? And, importantly, has *daylight* been shown to be beneficial (which would be good for TDDs) or is it just daylight*ing* through windows?

2.4 The Evidence: Daylight[ing] and People

This section is broken up into four parts, each covering one of the key facets examined by research into the human response to lighting. The four categories are:

1. People's preferences for and perceptions of lighting
2. Effects on mood
3. Effects on health
4. Effects on performance and productivity

2.4.1 Preferences and perceptions of lighting and the environment

It is clear that aspects of lighting can change how people perceive the lighting and the environment (Boyce et al., 2003). These aspects include the light level (Banu, 2007; Baron et al., 1992; Boyce & Cuttle, 1990; Tenner, 2003), the colour (Banu, 2007; Baron et al., 1992; Fotios & Gado, 2005; Knez, 2001), and the distribution of light (Hendrick et al., 1977; Tenner, 2003).

It may be expected that daylighting would affect people's perception of the lighting and environment, as it differs significantly from typical electric lighting. The question is whether or not the daylit rooms are responded to better than non-daylit rooms?

2.4.1.1 Daylighting and windows are much preferred

It is well established that people generally prefer daylighting to artificial lighting (Boyce et al., 2003; Collins, 1976). Surveys done in the UK, the USA, Canada, and New Zealand have consistently found a preference for daylight and studies have also found that people prefer to sit by windows (Boyce et al., 2003; Kilic & Hasirci, 2011; Shemirani et al., 2011).

But why do people prefer daylighting? A number of reasons have been suggested. Boyce et al. (2003) cite surveys of office workers that found that the most important aspects of windows are a view out, and the provision of daylight. Begemann et al. (1997) found that people generally prefer to follow the daylight cycle rather than having a constant light level. Based on a number of surveys, Cuttle (2002) suggested that it may be because people believe that artificial lighting is bad for one's health (Boyce et al., 2003). The findings of Young & Berry (1979), however, suggest that having a view is very important. They found that an artificial window with a fake "view", but no daylight, was liked almost as much as a real window. A number of other studies have also found that daylight is less important to people than view (Collins, 1976). This suggests that the view is far more important than daylight through windows.

2.4.1.2 People appreciate well daylit spaces, and dislike windowless spaces

There are also many examples of people responding positively to well-daylit spaces, and specifically noting daylight as an attractive quality (Edwards & Torcellini, 2002). A field study of a university library found that the amount of daylight was positively correlated with how comfortable users found the library (Kilic & Hasirci, 2011). This is, however, dependent upon the daylighting being well implemented. Poor daylighting that creates glare and overheating will not be positively responded to (Boyce et al., 2003; Edwards & Torcellini, 2002). It should also be noted that people's positive responses to daylit buildings mention view as much as daylight (Edwards & Torcellini, 2002).

Similarly, surveys have found that people dislike windowless offices, and often believe that the lack of windows has negative effects on them (Collins, 1976). Reports from factories with blacked-out windows tell of workers breaking the windows to get access to the outdoors (Sundstrom, 1986). The lack of daylight however is only one reason for this, with other complaints being the lack of view, the need for a connection to the outside, and desire for fresh air (Collins, 1976; Sundstrom, 1986).

2.4.1.3 The mere presence of windows can improve people's perceptions of the qualities of the space.

As mentioned previously, Carter & Marwaee (2009) found that people were more satisfied with the lighting when they had windows — even if the windows were not appreciably affecting the light levels at their location. Similar links have also been found in other studies (Boubekri, 1995; Charles & Veitch, 2002).

Studies have also found that glare may be better tolerated from daylight than from electric light (Galasiu & Veitch, 2006). In one study people were even willing to ignore glare when there was a good view available (Galasiu & Veitch, 2006).

2.4.1.4 But daylight is not necessarily very important

Heerwagen & Heerwagen (1986) had people rank how important various environmental factors were to providing a comfortable environment, and found that daylight is not considered that important: out of 20 environmental variables, having daylight was ranked 19th, behind factors such as being able to personalise one's workspace, and "a colourful interior". This suggests that daylight, while appreciated, is more of a luxury item. This interpretation is supported by the fact that people will get rid of the daylight if it is causing problems such as glare or if they want privacy (Boyce et al., 2003).

This argument would seem to conflict with the reports of people's dislike of windowless spaces. However, windows with blinds pulled generally still provide *some* daylight and connection to the outdoors, and having windows with blinds provides people with the *option* of daylight and view — and people appreciate control over their environment (Boerstra et al., 2013). Thus, windows with blinds pulled could be significantly better than no windows.

2.4.1.5 Overall Summary

Overall, the evidence indicates that:

1) People *want* daylighting and windows, and they respond positively to well daylit spaces better than windowless spaces.
2) The mere presence of windows may improve a space, and people may be better disposed to, and more forgiving of, daylight than artificial light.
3) It is important to ensure that daylighting does not cause problems such as glare or overheating.
4) It is however difficult to disentangle the positive responses to daylighting from the properties of windows. While people appreciate daylight, they also like views, and it is difficult to say how well a non-window daylighting system like TDD would work. Some research suggests that view may be the most important factor.
5) Furthermore, the overall importance of daylighting for environmental comfort and satisfaction is questionable.

2.4.2 Effects on mood

A number of studies have shown that lighting *can* influence people's mood (Boyce et al., 2003; Knez & Enmarker, 1998; Knez & Kers, 2000; Knez, 1995; McCloughan et al., 1999; Plitnick et al., 2010; Viola et al., 2008). Anecdotal evidence also shows that people *think* that daylighting improves their mood, and helps them feel better (Edwards & Torcellini, 2002). So, how strong is the evidence that daylighting can improve people's mood?

2.4.2.1 Correlations between daylighting and mood/satisfaction

A number of studies have reported possible links between daylighting and mood.

Leather et al. (1998) found that greater sunlight penetration was associated with greater job satisfaction and well-being, and less intention to quit, in office workers. They also reported positive effects of having a more natural view.

Leaman & Bordass (2000, cited in Boyce et al., 2003) similarly found a positive correlation between mean satisfaction and the proportion of building occupants seated by a window.

A more controlled study of people working in a room with a large window found a small decrease in negative mood over 20 minutes during the day, but no change at night, suggesting that daylight may help mood (Dasgupta (2003, in Boyce et al., 2003).

Additionally, studies of people working in windowless offices have found them to be less positive and satisfied, and to be less engaged in their work (Finnegan & Solomon, 1981). People have also reported that the lack of windows makes it dull and makes them feel depressed and tense (Ruys, 1970; Sundstrom, 1986). Blacked-out factories in the 1940s provide anecdotal accounts of increased irritation and friction between staff that were associated with the lack of daylight (Sundstrom, 1986).

2.4.2.2 Daylighting may affect health, and thus mood

Suppositions about daylighting's effect on mood are also indirectly supported by studies finding links between windows and health effects such as headaches and sleep quality (Aries, 2005; Boyce et al., 2003; Çakir & Çakir, 1998). As Boyce et al. point out: headaches are unlikely to be good for people's mood.

2.4.2.3 Circadian activating light may boost mood and alertness

One reason for daylight effects is that it has significant levels of light in the blue spectrum that the human circadian system is most sensitive to (Figueiro et al., 2011). By stimulating the circadian system, daylight could affect people's alertness and mood (Figueiro et al., 2011; Webb, 2006).

This hypothesis is supported by recent studies that looked at the effects of blue-enriched white light (17000K) compared to typical electric lighting (Mills et al., 2007; Rautkylä et al., 2010; Viola et al., 2008). Exposure to the blue-enriched light was found to improve alertness and reduce fatigue in office workers (Mills et al., 2007; Viola et al., 2008) and students in lectures (Rautkylä et al., 2010). It was also found to improve positive mood and irritability (Viola et al., 2008). Similarly, Lehrl et al., (2007) found that alertness was higher under blue light than under yellow or white.

While such findings support the case that daylight may have beneficial properties, they also suggest that electric lighting with more blue spectrum light can provide the same kinds of effects.

2.4.2.4 People are more positive on sunny days

The idea of daylight being good for people's mood is also loosely supported by other psychological studies. Sunlight and sunny days are known to induce and encourage good moods (Cohen, 2011; Cunningham, 1979).

2.4.2.5 However, studies into mood are unreliable, and reveal complications

While the research discussed above may make a good case for daylighting, other research is less positive. Indeed, as discussed by Boyce et al. (2003), research into lighting and mood has painted a sometimes confusing picture. The mechanisms involved are complicated, findings are unreliable, factors like gender and age can affect responses, and mood is influenced by a wide range of factors that make it difficult to predict the effects of any one factor (such as lighting) upon it (Boyce et al., 2003).

Boubekri et al. (1991) looked at the effects of window area and sunlight penetration on affective state. They found that sunlight penetration of between 15-25% of the floor area could improve feelings of relaxation, but *not* excitement or satisfaction — and only if the subject was sitting sideways to the window. If the subject was facing the window, there was no effect.

Boyce et al. (2003), in discussing Hartleb (1989) and Gutkowski (1992), note that the absence of any apparent effects of windows on mood in the studies could be due to only measuring it once, with any effects potentially masked by the individual differences in initial mood.

Stone & Irvine (1993) had students carry out various computational and managerial tasks in rooms with and without windows. They did not find any significant effects of windows on mood or performance. Indeed, they found that students felt slightly more confident with*out* the windows. This lack of effect on mood was also replicated in a later study (Stone, 1998).

The complexity and inconsistency of mood research is shown well in studies of the effects of the colour of light. Knez (1995) found that females had lower negative mood under warm-white (3000K) lighting than cool-white (4000K), while for males it was the opposite. Note that the first experiment only found effects on *negative* mood, while the second only found effects on *positive* mood. Knez & Kers (2000) also found effects of colour temperature on *negative* mood, as well as getting different results for older and younger subjects.

However, Knez & Enmarker (1998) had opposite results to Knez (1995), finding that females were better under cool-white lighting, while males were better under warm-white lighting. A possible explanation for this difference was that the ages of the subjects had a very wide range and was uncontrolled, which could have affected the results (Knez, 2001).

A similar study by McCloughan et al., (1999) supports Knez's (1995) results for females, but found no significant effect of colour temperature on mood for males.

It may be noted here that these findings suggesting positive effects of warm-white lighting would seem to contradict and weaken the conclusions drawn by the previously discussed studies suggesting positive effects of higher colour temperature (blue-spectrum lighting). At the very least, they suggest that the issue may be complicated.

The findings about the effects of colour temperature on mood are also questionable. In another study, Knez (2001) found *no* significant effect of colour temperature on mood and questioned the use of the PANAS tool for measuring mood in the previous studies.

A study comparing "full-spectrum" (>5000K) and cool-white lighting suggested that people's beliefs about lighting could influence their responses — that people might have better mood because they believed that the lighting should improve their mood (Veitch et al., 1991). A later replication however failed to reproduce the demand effects — and didn't find any effect of the different lighting on mood or performance (Veitch, 1997).

Boray et al. (1989) also found no effect of the colour temperature on mood when comparing warm-white (3000K), cool-white (4000K) and "full-spectrum" fluorescent lighting.

It's clear that the effects of factors such as light spectrum and windows on mood are complex, and any effects on mood may not be reliable, or may involve interactions with other factors.

2.4.2.6 Daylighting may only provide reliable effects over the long term

There are a number of studies that say that lighting, and daylighting, can positively affect people's disposition. There are also a number of studies that cast doubt on such conclusions, raising complications and concerns about reliability.

However, there is one factor that seems to critically differentiate the two groups of studies. Virtually all of the studies that provide support for the benefits of lighting and daylighting are *field studies*, while the ones that failed to find positive effects, or showed great inconsistency, are all shorter *laboratory studies*. A possible explanation could be that lighting has much more influence on people over long term exposures — such as in office workers that experience it day after day for hours on end. People that are exposed once for a relatively short period may be much less affected, explaining why laboratory studies show inconsistent and unreliable results.

2.4.2.7 Overall summary

Overall, it can be said that:

1) There are good grounds to believe that daylighting *can* improve people's mood.
2) Blue-spectrum light may be the best kind of light for improving mood and alertness, due to it being best at stimulating the circadian system. As daylight has high levels of blue-spectrum light, it suggests that it may be good for people.
3) Views may be an important factor for benefits from windows.
4) Daylighting may require long term exposure to ensure positive effects. Laboratory studies with short exposures have been unreliable.
5) Mood may be influenced by a large number of factors that make it complicated to study and make effects of daylighting unreliable.

2.4.3 Effects on health

2.4.3.1 Links between daylighting and better health in workers

There are a number of studies that have found evidence of positive links between daylighting and health in workers.

Surveys of German office workers found that workers closer to windows had less health problems, stress, fatigue, and eye discomfort (Çakir & Çakir, 1998).

A study of office workers' health found that those on the lower floors suffered more frequent headaches, possibly because they received less daylight (Wilkins et al., 1989). Of note is that the view (of other office buildings) was the same on all floors.

Wotton & Barkow (1983, cited in Galasiu & Veitch, 2006) found that workers in office buildings with little (11%) glazing suffered more headaches than average.

A European study found that workers under combined artificial light and daylight reported lower stress levels than ones under only artificial light during May, when there was significant daylight penetration, and not in January, when there was little daylight (van Bommel, 2006).

Other studies suggest that the high levels of vertical illumination provided by windows may benefit health. Aries (2005) found that higher vertical illuminances were correlated with lower fatigue and better sleep quality. On a similar basis, another study found that the more satisfied people are with their working environments - which can be enhanced with windows (see section 2.4.1) - the less discomfort they report, and linked it to sleep quality (Aries et al., 2010).

Possible health effects have also been found in factory workers. Russian and Czechoslovakian studies have found that workers in windowless factories were more prone to sickness, and had higher absenteeism rates (reported by Plant, 1970, in Edwards & Torcellini, 2002).

These studies suggest not just that daylighting is good for people, but also that the level of daylight is important for people's health.

2.4.3.2 Windows may help hospital patients get better

There are also a number of healthcare related studies that have found benefits from windows and daylighting.

Joarder & Price (2012) found a correlation suggesting that 100lx of daylight in the hospital room could reduce coronary patient's length of stay by an average of 7.3 hours after controlling for other factors such as view. In another study, spinal surgery patients in sunnier, brighter rooms experienced lower stress and needed less pain relief (Walch et al., 2005). Similarly, a study of patients suffering from myocardial infarction found that patients in sunny, south-facing rooms had lower mortality rates compared to those in north-facing rooms (Beauchemin & Hays, 1998). They also found that patients in the sunny rooms had a shorter length of stay, though the difference was only significant for females (Beauchemin & Hays, 1998).

An older study by Keep et al. (1980) studied patients in an Intensive Therapy Unit. Patients in the windowless unit had more than twice as many hallucinations and delusions compared to those with windows — which had translucent glazing — so, as with Joarder & Price (2012), the benefit was not a function of view.

Perhaps reflecting mood effects, a study of depressed patients in a psychiatric unit found that patients in sunny rooms had a significantly shorter length of stay than patients in non-sunny rooms (Beauchemin & Hays, 1996), again suggesting effects of daylight level, or at least light level.

Some studies show effects that vary depending on certain factors. Rashid & Zimring (2008) reported a study that found that bipolar patients exposed to direct sunlight in the morning had shorter stays — but that the same was not true for unipolar patients. A study of *critically* ill patients with severe brain injuries found *no* effects of window rooms on patient recovery or outcomes (Wunsch et al., 2011). This suggests that effects may depend on the nature of the health problems.

It should also be noted that, while the above studies provide strong evidence for the benefits of daylighting, even accounting for view, there are also studies suggesting view is still important: Ulrich (1984) found that surgical patients in rooms with natural views had fewer problems, and left the hospital sooner, than patients in rooms with views of a brick wall. Another study found that prisoners whose windows had rural views had significantly fewer health problems than those whose view was of the prison yard (E. O. Moore, 1981).

2.4.3.3 Effects may be due to increased light levels

The various health effects linked to increased levels of daylight do not, necessarily, mean that daylight is actually beneficial. The benefits may simply result from the increase in illuminance rather than daylight specifically. If so, the same benefits could be achieved by simply providing more electric lighting. High illuminances are activating to the circadian system and are known to be viable means of treating certain conditions, such as depression and seasonal affective disorder (Boyce et al., 2003; Walch et al., 2005).

This would not necessarily be a problem for windows — after all, windows *do* typically provide significantly higher light levels than typical electric lighting levels (e.g. Heschong Mahone Group, 1999), so unless people start putting in much more electric lighting they would still be superior. Moreover, they would still provide views. However TDDs do not have such advantages (section 2.5).

2.4.3.4 Circadian activating light may improve health and well-being

There is, however, reason to believe that daylight itself may be beneficial. As has been previously discussed, daylight may affect people by stimulating the human circadian system with its higher levels of blue-spectrum light (Figueiro et al., 2011; Webb, 2006; van Bommel, 2006). Studies provide some support for the idea that blue spectrum light may be healthier for people: Viola et al. (2008) found that blue enriched white light improved eye discomfort and subjective sleep quality in office workers, while Mills et al. (2007) found that it significantly improved ratings of mental health and vitality in office workers over a 7 week period.

2.4.3.5 Daylighting may cause health problems via glare

However, daylighting must be well implemented to be beneficial. Poor daylight that causes glare may *adversely* affect people, and cause problems such as increased stress (Boyce et al., 2003). In the previously mentioned study by Wotton & Barkow (1983) workers in the buildings with high (68%) glazing reported more eyestrain, and had twice as much absenteeism. For that reason, they suggested that both low levels *and* high levels of daylight could be bad for people (Galasiu & Veitch, 2006).

2.4.3.6 There is little evidence for effects in schools

Research into health effects of daylighting in schools is less positive. Larson et al. (1965) studied the effects of windows on school children over several years. When the classroom windows were removed in the second year, the younger children showed significantly higher absenteeism rates, but the older children did not. Küller & Lindsten (1992) studied four classrooms with combinations of different electric lighting and daylighting. Students in one of the windowless classrooms showed a different pattern in levels of the stress hormone, cortisol, having the lowest cortisol concentrations in February rather than December. Cortisol was related to the ability to concentrate and to sociability. This suggests a *possible* effect of daylighting on hormone patterns. However, the effect was only found in one of the two windowless classrooms, and they only found a change in the annual pattern of cortisol — the study did not suggest that average health or ability to concentrate differed across the classes. While it does suggest possibilities worthy of further research, it does not actually demonstrate that daylighting is better or worse than electric lighting for student's health.

The much larger study by the Heschong Mahone Group (2003a) did not find any correlation between environmental characteristics and absenteeism, raising doubts about health effects in schools. However they did suggest that absenteeism might not be a very accurate measure of health in schools, due to students frequently being absent for non-health reasons.

2.4.3.7 Overall summary

Overall we can say:

1) There is good evidence that good daylighting via windows may have positive health effects on people.
2) Both the level of daylight and the view seem to be important.
3) There is reason to believe that blue-spectrum light, and thus daylight itself, may be beneficial.
4) As always, it is important to avoid poor daylighting that causes problems like glare, as it may have negative effects.
5) There is substantial evidence of beneficial health effects for office workers and hospital patients, but not for effects on school students.

2.4.4 Effects on performance and productivity

2.4.4.1 Daylighting may improve performance in schools

The strongest evidence for performance effects of daylighting comes from a set of studies done by the Heschong Mahone Group (Heschong Mahone Group, 1999a, 2003a). In the first study (Heschong Mahone Group, 1999a), analysis of primary school test scores from roughly 20,000 students over a year showed that the students in the classrooms with the most daylight had 20-26% higher improvement in test scores than the ones with the least daylight. Additionally, classrooms with the most daylight had 7-18% higher average test scores, and skylights that provided *diffuse* light to the classroom were associated with 19-20% better test improvement. The researchers noted that the effect of daylight was greater than the effect of larger windows, suggesting that daylight itself could be providing positive benefits — not just view.

Caution is advised here. The effects reported above suggest that daylighting is *extremely* important. Closer examination of the statistics reveals a less positive picture, as discussed by Boyce (2004) in a review of the studies. The regression equations only manage to explain 26% of the variance in the data, and the daylight variable explains only 0.3%. This means that most of the variance in the test scores cannot be explained by the model, and that daylight had a very small effect (Boyce et al., 2003). Moreover, the significance of a variable that explains so little of the variance is heavily dependent on the other factors in the equation — and it could easily lose significance if the other factors change (Boyce, 2004). It has been pointed out, however, that the small effect size of daylight is still comparable to other factors that are widely felt to be important, such as participation in a gifted and talented program, and absenteeism (Heschong et al., 2002).

The second study (Heschong Mahone Group, 2003a) was an attempt to replicate the first in another district while controlling for more factors. However, they failed to replicate the previous results. Instead they found that the best performance was in both the classrooms with high daylight and those with low daylight. They linked this to interference with other environmental factors — particularly noise. Classrooms with high levels of glazing had higher reverberation times, meaning that the poor acoustics could make it difficult for the children to hear. Differences in school designs meant that they had problems with teaching assistants carrying out tutorials in the back of the rooms — increasing noise levels — while in the schools with low levels of daylighting they would typically run the tutorials in common areas outside the classrooms. More windows also meant more *open* windows — which could be a problem in that district because the outdoor air quality was poor, and crowding meant that schools run staggered lunch breaks, so there would be students outside having lunch while others would be in class.

Older studies have also been much less positive, finding little evidence of benefits from windows in schools (Collins, 1976). Demos et al. (1967, cited in Collins, 1976; Wu & Ng, 2003) studied the performance of students in two classrooms — one with windows, and one without — over two years. They found no significant differences in performance — though it should be noted that such a limited sample would not be able to deal with possible confounding factors such as differences in the class groups and teachers. A more powerful study was carried out by Larson et al. (1965), who studied a primary school over three years, with the windows being removed in the second year. Again, no effect on performance was found.

Overall, the benefits of daylight on education are still inconclusive, and need more study — although it is not unreasonable to suggest that well-designed daylighting could have positive effects (Boyce et al., 2003). It is also apparent that studying this is very difficult.

2.4.4.2 Daylighting may improve performance in offices

There is also some evidence suggesting that daylighting may improve performance in offices.

The Heschong Mahone Group (2003c) carried out a study in a call centre, measuring worker productivity, and another study running a number of cognitive tests on some 200 office workers. The strongest effects of windows were those of view. They found that the call centre workers with the best views performed 6-12% faster compared to those with no views. They also found that the office workers with the best views performed 10-25% better on the cognitive tests. Better health was found to correlate with good views, while lack of views was connected to higher fatigue.

Daylight levels, however, had much less of an effect, and were not consistent, showing effects on only the two of the cognitive tests. The most notable effect was a correlation with performance on a working memory test, predicting that an increase from ~10lx to ~215lx of daylight would improve performance by 13%.

As with the school studies, however, the regression models used to calculate these effects show that the environmental factors have only very small effects, with the window factors generally explaining less than 1% of the variance — suggesting that their effects are likely to be unreliable (Boyce, 2004). As pointed out in the report, though, even very small effects in productivity may be valuable (Heschong Mahone Group, 2003b).

From field surveys of office buildings, Leaman & Bordass (2000, cited in Boyce et al., 2003) reported a negative correlation between building depth and perceived productivity — the deeper the building, the lower the reported productivity. As discussed by Boyce et al. (2003), there are a number of possible reasons for such an effect such as the fact that larger buildings do tend to require more artificial environmental control, while smaller buildings may find it easier to make use of things like natural ventilation and daylighting. However, combined with the fact that they also found a positive relationship between the proportion of people sitting next to a window and mean satisfaction, it *suggests* that windows may have a positive impact on productivity.

(Hedge, 1994, in Boyce et al., 2003) assessed performance of subjects on a computer-based clerical task in a room with and without windows, and found a small improvement in task performance with windows. However, Santamaria & Bennett (1980) had subjects carry out several tasks under daylight and fluorescent light with identical light levels and distribution, but only found one difference — that proofreading was done 5% faster under daylight. This may be explained by effects of flicker from the electric lighting (Boyce et al., 2003) — something which is far less of a problem for modern electronic ballasts (Wilkins et al., 1989). Several other studies looking at performance on computational and managerial tasks have consistently found no effect of windows on performance (Stone & Irvine, 1993, 1994; Stone, 1998).

Generally the research into the possible benefits of daylighting on office workers is suggestive, but the evidence in favour is not strong. It may also be suggested that any effects may be more due to view than to daylight.

2.4.4.3 Daylighting may improve retail sales

Other studies have suggested that the presence of daylighting — in particular skylights — may positively impact retail sales.

The Heschong Mahone Group carried out two such studies (1999b, 2003a). In the first study, they analysed sales data from 108 stores looking for relationships between sales and factors such as size, hours open, and the presence of skylights. The resulting regression model explained 58% of the variance in the data. The presence of skylights explained 4% of the variation, making it a small — but significant — effect (Boyce, 2004). Moreover, its predicted effect on sales was high, suggesting that adding skylights to a store could improve sales by an average of 40%.

The second study was an attempted replication. However, they failed to replicate the previous study's findings, with daylighting having a much weaker effect. Instead, they found an interaction between daylight and *parking*, with daylighting only managing to achieve strong effects if there was plenty of parking. Indeed, when there was below average parking, daylighting appeared to *reduce* sales. Daylighting did, however, have as much explanatory power as other factors believed to be important, such as number of competitors and neighbourhood demographics (Heschong Mahone Group, 2003c).

Overall, they suggest that daylighting *may* on average provide at least a small boost in sales, but that effects may be moderated by other factors, such as whether or not there is sufficient parking to handle increased demand.

Some further support for the notion is provided by a case study (Romm & Browning, 1994). Monitoring of sales in a partially skylit Wal-Mart revealed that sales in the daylit half of the store were significantly higher than in the other half or those in the same departments in other stores (Romm & Browning, 1994). While one case study is not conclusive, it does provide support to the findings of the Heschong Mahone studies.

2.4.4.4 Circadian activating light may enhance performance

As has been previously discussed, daylight may affect people by stimulating the human circadian system with its higher levels of blue-spectrum light (Figueiro et al., 2011; Webb, 2006; van Bommel, 2006).

Studies of the effects of blue-enriched white light (17000K) support the hypothesis. Mills et al. (2007) found that workers under blue-enriched white light showed significantly improved subjective work performance, alertness, and fatigue. Similarly, Viola et al. (2008) found that blue-enriched white light could significantly improve ratings of subjective performance, concentration, alertness, and fatigue in office workers.

Experimental studies also show some support for the hypothesis. Lehrl et al. (2007) exposed subjects to different colours of light for brief periods, and found that alertness and speed of information processing was significantly higher under blue light than under yellow or white light. Hoonhout et al. (2009) found that subjects completed a proof-reading task significantly faster under blue light than under red light (~10min vs. 11min). However, there was no significant difference in the number of errors.

Other studies, however, have found better performance under warm-white (3000K) lighting compared to cool-white (4000K and 5500K) on short-term memory and problem solving tests (Hygge & Knez, 2001; Knez, 1995, 2001), which would seem to contradict the "blue light is better" hypothesis. This may suggest that different tasks may be affected by lighting differently, or that the influence lighting has on people is complicated and effects may not be reliable.

2.4.4.5 Bright light may enhance performance

There is also evidence that higher illuminances may improve people's performance.

Buchanan et al. (1991) studied the error rate in prescription-dispensing, and found that increasing illuminance from 485 to 1570 lux reduced the error rate from 3.9% to 2.6%. It is also known that bright light can help improve performance of night workers (Boyce et al., 2003; Kretschmer et al., 2011; Webb, 2006).

Indirect support is provided by studies suggesting that bright light can boost alertness: Küller & Wetterberg (1993) found that subjects were less sleepy under 1700lx than 450lx. Noguchi et al. (2004) found that bright light in the office in the morning and after lunch could improve alertness and mood.

This may provide at least a partial explanation for possible daylighting effects, as daylighting often provides greatly elevated light levels (Webb, 2006).

2.4.4.6 Attractive lighting that helps people feel better *might* enhance performance

While attempts to directly measure effects of lighting on performance have had mixed success, it may be argued that positive effects on mood and environmental satisfaction *indirectly* provide evidence for positive effects on productivity.

There is considerable evidence that positive mood can positively impact people's performance and behaviour. In a meta-analysis of the literature, Isen (2001) concluded that positive mood could encourage people to be more generous and kind, be more socially responsible, and better consider other people's point of view — leading to better outcomes in negotiations. People in a good mood may also be better at solving problems, being more flexible, creative, and willing to consider multiple ideas, while also being more thorough and analysing information more efficiently. All this would clearly be beneficial — particularly in jobs where such skills are used. There are also studies that have found direct impacts on performance — for example, Miner & Glomb (2010) found that customer service workers completed calls roughly 5% faster when they were in a positive mood, with no trade-offs in service quality.

Lighting may be used to help promote positive mood in people, and in turn encourage more pro-social behaviour (e.g. Baron et al., 1992). Veitch et al. (2011) developed a model showing that lighting appraisal could predict room attractiveness, workplace satisfaction, pleasure, and work engagement. Given that people prefer and respond well to daylighting (section 2.4.1), this shows how it could positively impact productivity.

2.4.4.7 Effects may be task-specific

Effects may depend on the task at hand. Often studies looking at performance only manage to find effects on *some* of the performance tests they run (e.g. Heschong Mahone Group, 2003b; Knez, 1995, 2001; Santamaria & Bennett, 1980). This is not too surprising —the different tests *are* designed to test different skills — however it does highlight both the difficulty in measuring performance, and the fact that benefits of daylighting may be limited to specific situations or tasks rather than it being generally "better".

2.4.4.8 Bad daylighting may reduce performance

As discussed earlier, positive effects of daylighting are reliant upon it being implemented well. Daylighting that causes glare is likely to *impair* performance (Boyce et al., 2003). In the Heschong Mahone studies (1999a, 2003b), glare from skylights that provided large amounts of *direct* sunlight was associated with a 21% *reduction* in performance on reading tests. In the study of office workers, high glare potential from the windows was associated with performance reductions of 15-21%.

2.4.4.9 In some cases views may be distracting

There are some studies suggesting that views can negatively affect people's performance. Demos et al. (1967) (cited in Collins, 1976) noted that some students preferred windowless classrooms because of the absence of distractions. Similarly, Larson et al. (1965, in Collins, 1976) found that the teachers appreciated a change to windowless classrooms, as they felt that there were less distractions. When the windows were returned, they complained that the students were distracted more. Edwards & Torcellini (2002) also references anecdotal evidence from students and teachers in universities suggesting that windows can be distracting.

The Heschong Mahone Group study (2003b) also suggests that distractions through windows can be a problem. They identified possible negative effects from windows that they suggested could be partially due to students being distracted by outside noise and activity.

TDDs would be good daylighting from this point of view — providing good daylight without distractions.

It should be noted, though, that people also complained about factors such as temperature swings, poor ventilation, and glare, suggesting that objections to windows may have been exacerbated by poor design (Collins, 1976).

2.4.4.10 Overall summary

So, overall we can say:

1) There is reason to believe that daylighting *may* be able to improve performance in schools and offices, however research is still not conclusive.
2) Skylights may be able to increase retail sales.
3) Effects of the environment on people are complicated, with many different factors involved. Desired effects may not be reliable simply because they are also being affected by other, unconsidered factors. Measuring performance is difficult.
4) It is important to avoid problems such as glare, as they can have negative effects.
5) View is also important — though in some cases it may be distracting.
6) Studies also suggest that blue-spectrum light may be good for alertness and performance — suggesting that daylight itself may be good.
7) Studies finding positive effects on mood and lighting appeal may indirectly provide evidence for effects on productivity.

2.4.5 Overall

1) People *want* daylighting and windows. The presence of these can help people perceive a space more positively.
2) There is reasonable evidence that daylighting - and possibly daylight itself — can help improve people's mood. However, mood is influenced by a large range of factors and effects may not be reliable or consistent.
3) There is strong evidence that daylighting — and possibly daylight itself - may have positive health effects on workers and hospital patients. However, evidence for health effects in school students is weak.
4) There is reason to believe that daylighting — and possibly daylight itself — *may* be able to improve performance in schools and offices, but research is still not conclusive. There is better evidence that skylights may be able to enhance retail sales.
5) Field studies with longer term exposures seem to find the most positive effects, while laboratory studies with much shorter exposures are generally less positive and less likely to find effects. This suggests that daylighting needs long term exposure to reliably have significant effects on people — possibly by affecting their health.
6) Often it is difficult to separate out benefits of daylighting from other aspects of windows such as views. The literature indicates that both daylight and views are important. Some studies suggest that the view is most important. Hence the evidence for daylight itself being beneficial is weaker.
7) Any positive effects are reliant upon the daylighting being well implemented and not causing problems such as glare. If it does cause problems, then it may negatively impact people's health and productivity.

So, daylighting *does* make spaces more appealing, *good* daylighting is probably better for mood and health, and while there is reason to believe it may affect performance, results are still inconclusive. Also, views are at least as important, if not more important, than daylight.

2.5 Why daylighting being beneficial doesn't mean TDDs are

Now that we have examined what the possible effects of traditional daylighting are, we need to ask how they may apply to Tubular Daylighting Devices. To do this, we look at how the possible positive aspects of windows apply to TDDs.

2.5.1 Mechanisms for traditional daylighting to provide benefits

There are a number of possible reasons that may allow traditional daylighting to provide benefits beyond that of standard electric lighting:

1) **Spectrum** — as has already been discussed, it is argued that the spectrum of daylight may be better than standard electric light — in particular, that its high levels of blue-spectrum light may be better at stimulating the circadian system (Figueiro et al., 2011; Webb, 2006).
2) **View** — another important aspect of windows is the view they provide. Evidence suggests that views, particularly natural ones, are appreciated by people and may have beneficial effects (e.g. Moore, 1981; Ulrich, 1984; Young & Berry, 1979).
3) **Higher light level** — windows and skylights often provide light levels that are much higher than those provided by artificial lighting (Webb, 2006), which may account for many of their benefits (Heschong Mahone Group, 1999a).
4) **Better light distribution** — typical electric lighting practice involves directing much of the light downwards, towards the horizontal plane (Boyce et al., 2003). In contrast, windows provide much more vertical illuminance, providing much more light to the eye, and thus provide greater circadian stimulation (Boyce et al., 2003).
5) **Flicker** — flicker in electric lighting has been found to negatively affect people (Wilkins et al., 1989), and so daylight's *lack* of flicker may be a benefit. Note, however, that modern electronic ballasts have much less of a problem in this regard (Wilkins et al., 1989), so it may be more of an issue to consider when looking at older studies and may be less of a benefit for daylighting in contemporary buildings.
6) **Psychological association** — it is also possible that windows may provide benefits simply because people appreciate having daylight and views and *think* they're good, and so having them makes them feel better. This is supported by the studies finding that the mere presence of windows could improve people's perceptions of spaces. Demand effects are a possibility for lighting (Veitch et al., 1991), and it has been suggested that people may like daylighting because they think fluorescent lighting is bad for them (Cuttle, 2002, cited in Boyce et al., 2003).

2.5.2 The much smaller set of mechanisms employed by TDDs

TDDs are in many ways more similar to electric lighting than to traditional daylighting. They are much closer to electric lighting in appearance than they are to windows. They do not provide a view. They also do not provide the very high light levels that windows do — indeed one of their selling points is their ability to manage and prevent excessive illumination. While their light distribution may differ from common electrical lighting, they are hardly similar to wall mounted windows — although the few studies finding positive effects from skylights suggest that providing daylight from the roof can still be beneficial.

Ultimately, TDDs are primarily reliant on *daylight itself* being inherently better than the spectrum of electric lighting if they are to be superior. They *do* also lack flicker problems — but as noted, this is significantly less of an issue for contemporary lighting. They could also benefit from psychological effects — however any effect would probably be weaker than for windows as TDDs lack important features such as views, and research suggests that people don't appreciate the daylight from TDDs as much as that from windows (Marwaee & Carter, 2006).

2.6 Conclusion: The case for TDDs

Current research, which is rather sparse, does not say whether or not TDDs are better for people than typical electric lighting. The evidence for daylighting shows that *windows* are generally preferable, and may improve mood and health, but that attempts to demonstrate performance effects have been inconclusive. The evidence is even weaker when extrapolated to TDDs, due to the significant differences between TDDs and traditional daylighting strategies.

However, while the evidence in favour of daylight itself being beneficial may be weaker, there is still reason to believe that it, and TDDs, *may* be "better" — with respect to creating a more pleasant, healthy, and productive environment — than typical electric lighting. The hypothesis just needs to be tested.

3 METHODOLOGY

This chapter covers the study methodology, with a focus on its development and the reasoning behind the decisions. Discussed are its concept, the environmental conditions used for testing, the limitations and issues and how they were addressed, the selection of the most appropriate tests, and the details of the tests the subjects did.

3.1 Basic concept: the opportunity

The methodology was developed from a simple proposal. Specifically, that Hometech would install TDDs in one of the computer rooms in the School of Architecture, and then we could carry out a before and after study on the students to see how they responded. This idea had a number of advantages:

1) **Large sample available** — several hundred students use the room for tutorials on a regular basis, providing a convenient source of subjects. Additionally, the rooms can support a substantial number of students (~40), so many people can be tested at once.
2) **Consistent use from the same groups** — the same groups of students use the room regularly throughout the course of the semester. This means that the same groups, and individuals, can be assessed both with electric lighting and with TDDs. This allows more sensitive within-group analysis that controls for individual differences.
3) **Windowless** — the computer rooms are windowless, making it easy to directly compare electric lighting and TDDs without the confounding factor of windows.
4) **Control groups available** — also valuable is the fact that there are other similar computer rooms that are also used for tutorials. By assessing students in them at the same time as the test room, we gain several advantages: control groups that can show us what the variation in response is without TDDs; larger samples of students under electric lighting; and the opportunity to compare small environmental differences between the computer rooms.
5) **Convenient to study** —the computer rooms at university are relatively convenient and easy to study.

3.2 Study parameters: the rooms

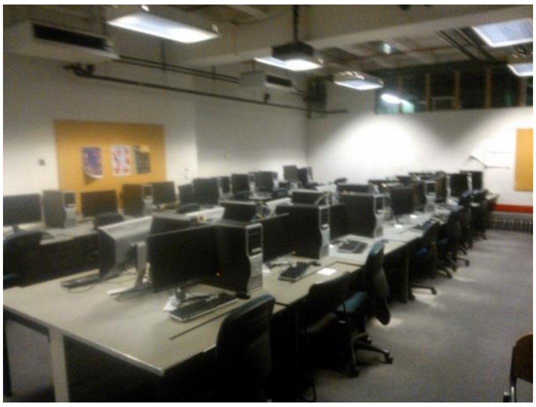
3.2.1 Visual environment

Note: plans and details may be found in Appendix B.

Test room: VS319 (Electric lighting)

Test room: VS319 (under TDDs)

Control: VS322 (Electric lighting)



Control: VS226 (Electric lighting)

Figure 3-1: Photos of the different rooms

3.2.1.1 The rooms are generally similar in basic design and appearance

	VS319 (Electric light)	**VS319 (TDDs)**	**VS322 (Electric light)**	**VS226 (Electric light)**
Horizontal illuminances on desks (lux)	460	400 Sunny: 520 Overcast: 190	310	330
Vertical illuminances on screens (lux)	140	190 Sunny: 250 Overcast: 90	110	210
Colour temperature	4000K	Daylight: 5000-10,000K	4000K	4000K

Table 3-1: Light levels measured in the different rooms (measurement details in Appendix B2). Daylight's colour varies depending on the weather (Veitch & McColl, 1994).

As shown above, the three computer rooms are similar in design: rows of computers, white walls, dark carpet, grey desks, and no external windows - though there are some internal ones (Figure 3-1). The TDDs were installed in VS319 before the semester started, and blacked out. Thus, they were present in the room while its electric lighting was assessed, and so the design of the room did not change between the electric lighting and TDD tests.

The lighting in all the rooms is laid out in rows running parallel to the computer, with the luminaires generally directing the light downwards towards the desks and casting pronounced shadows on the walls. The lamp colours are all the same. There is, however, some variation in light levels. Illuminances were measured at each computer on the desk and on the screen. VS319 has higher workplane illuminances, while VS226 has higher illuminances on the computer screens because its lights are running *between* the rows of computers, whereas in the other rooms the lights are aligned more *above* the computers.

The electric lighting in VS319 and VS226 consists of large suspended luminaires with "egg-crate" style diffusers, holding four fluorescent lamps each (Figure 3-2). The lighting in VS322 consists of suspended luminaires with parabolic reflectors and louvers, each containing two lamps (Figure 3-3).

Overall, the rooms are similar in appearance, and this supposition was confirmed by comparisons of people's perceptions of the three rooms (section 4.1.1).

Moreover, the differences between the rooms do not cause any problems with using VS226 and VS322 as controls because the key element in their use as controls is *consistency.* Specifically, that their lighting does not change during the study, so any variation in their results must be due to non-lighting factors. Also, differences can be useful — as discussed below.

Figure 3-2: Luminaire in VS226. Note the "egg-crate" style diffuser, and the lip around the fitting that suggests it should be recessed.

Figure 3-3: Luminaire in VS322.

3.2.1.2 Small differences are useful

Differences between the rooms are also useful, as they can help tell us what environmental factors may be influencing the results. In addition, having a degree of variation amongst the rooms can make the results more robust and capable of being extrapolated. If TDDs get better responses in the test room, then it could be because of something specific to that room, limiting the results. If, however, the TDDs get better responses than the electric lighting in *any* of the rooms, then the results would be stronger and more broadly applicable. Similarly, if TDDs were better than *some* rooms but not others, then the differences between them could help explain which aspects of the lighting are responsible for the different responses.

3.2.1.3 The electric lighting is fairly typical for the post-1980's

The lighting in the rooms was designed by BECA in the 1990's. According to them, the lighting layout is fairly normal, as is the use of 4000K lamps (Hirschberg, 2012). The light levels are normal for an office space, meeting the levels recommended by the building code (NZS 1680.1). The luminaires are ones commonly used in the 1990's: The "egg-crate" diffusers in the fittings in VS319 and VS226 have a sharp cut-off, and were used at that time because it was important to minimise glare on the CRT monitor screens. The louver fittings in VS322 are a type that started being used in the late 1980's, and were the most commonly used fitting until recent times. Indeed, they are still used, because they are efficient. They may be considered typical "post-1980's" lighting.

These days, as LCD computer screens are less vulnerable to glare problems, lighting designers often use more diffuse lighting, trying to get more light on the upper walls and ceilings to make more comfortable spaces. In this case, they would likely add an uplight component to the lighting to provide indirect light and reduce the shadowing (Hirschberg, 2012).

The lack of contemporary quality electric lighting is a limitation of the study. However, the lighting assessed includes some of the most common lighting, and some comparisons may be made with other studies that have assessed newer lighting.

3.2.1.4 TDD design and comparison

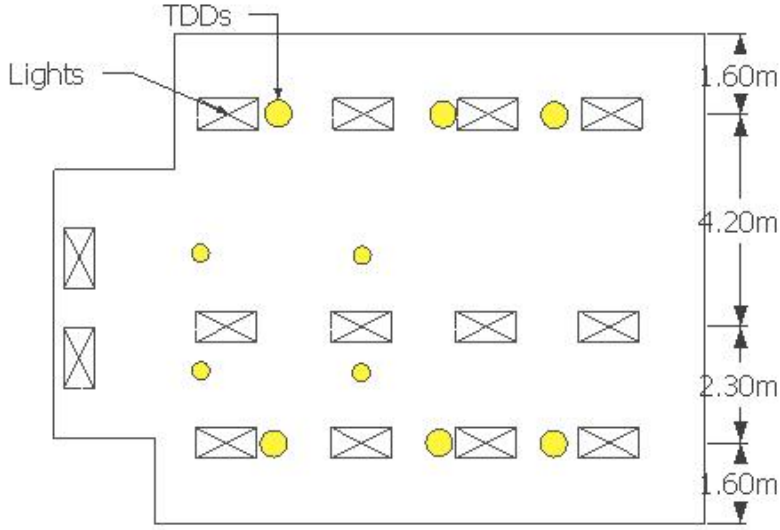


Figure 3-4: VS319 ceiling plan (1:150). All luminaires/TDDs are at the same height: 2.8m

The TDD layout was designed by Hometech to be similar to the existing lighting, and to provide an average workplane illuminance meeting the requirements of NZS 1680.1.

The basic layout and optimum placement was planned using IES and photometric files of the TDDs. However, to work around issues with the structure and HVAC units on the roof, the basic plan was adjusted using professional rules of thumb.

So, two rows of large 530mm wide Solatubes were run down the room in line with the electric lights — one row on each side (Figure 3-4). Four smaller 350mm tubes were placed in the middle-rear section of the room to provide more light there, while none were placed in the front of the room because they would have interfered with the projector screen.

The TDDs provide very different lighting. The most notable difference is a more diffuse light (Figure 3-1). The electric lighting casts sharp shadows on the walls, while the diffuse light of the TDDs does not. The light levels provided vary - on a sunny day the room is brighter overall, with similar horizontal illuminances to the electric lighting, and higher vertical illuminance. On overcast days, however, the light levels are lower (Table 3-1).

The daylight also has a cooler, more blue-white cast to it than the lamps.

3.2.2 Other environmental conditions

	VS319 (Electric light)	VS319 (TDS)	VS322 (Electric light)	VS226 (Electric light)
Temperature (°C)	21.7	21.3	21.9	22.5 - 24.8
CO_2 (ppm)	700	630	1200	830
Noise levels (dB(A))	55	59	61	55

Table 3-2: Average environmental conditions measured in the different rooms (measurement details in Appendix B2)

Other environmental factors are generally fairly typical for an educational/office environment, as discussed below.

Temperatures are mostly within the typical comfort range of 20-24°C (NZS 4243) with the exception of VS226 in the afternoon, where temperatures can rise up to 25°C.

CO_2 levels vary, with VS319 having the lowest levels. Both VS319 and VS226 keep CO_2 levels below the recommended limit of 1000ppm (NZS 4303). VS322 has CO_2 levels rise significantly above that, which might affect performance (Coley & Greeves, 2004).

Noise levels are higher than recommended levels of 45dB(A), but are within typical ranges for offices and schools (European Agency for Safety and Health at Work, 2005).

The possible impact of variance in these conditions is discussed in the analysis (4.2.1).

3.3 Limitations and issues

Building on the basic concept, the methodology was refined by assessing and addressing its limitations and experimental issues. These included:

1) The need to avoid negative responses to the tests
2) The varying times of the tutorial sessions
3) The unpredictability of the weather
4) Differences between the groups
5) The variability of the tasks students are doing in their tutorials
6) The limited number of days that tests could be run
7) Tutorial length limiting exposure time
8) Variation in light levels around the room
9) The representativeness of the sample
10) Uncertainty in participation rates

These are discussed below.

3.3.1.1 It is important to minimise the inconvenience of the tests

As the tests rely on voluntary participation from the subjects, it is important to avoid negative responses. This suggests several things:

1) It is necessary to avoid taking up too much time. This is important both so as not to annoy the subjects, but also to avoid harming their education by using up too much class time. Five minutes is suggested as a reasonable limit in a 1 hour tutorial. Slightly longer tests (e.g. <10 minutes) may be able to be used *if* only a few tests are run.
2) To maintain interest, a variety of tests should be used. Subjects are likely to get bored and stop if they are asked to do the same test again and again. By using multiple different tests, and rotating between them, interest may be maintained longer.
3) It is important to avoid times of high stress when students are very busy and won't want to spare the time. No tests should be carried out in the same week as assignments are due.

3.3.1.2 Tutorials being at different times of day makes the results vulnerable to confounding circadian effects

The roughly one hour long tutorial sessions where the students were tested were run at set times in the day:

- In VS319 (the test room) three tutorials were run at 11:30am, 12:40pm, and 1:40pm.
- In VS322 (control room) two tutorials were run at 9:30am and 10:30am.
- VS226 (control room) had tutorials in both the morning and afternoon, at 9:30am and 1:40pm.

Due to circadian rhythms, people's alertness is different at different times of day (Begemann et al., 1997). Thus, the control groups may differ from the test group in aspects such as alertness or performance simply because they are being measured at different times of day rather than because of differences in lighting.

The problem is addressed by assessing both the control rooms and the test room under electric lighting at the same time before testing the TDDs. As the electric lighting is generally similar across the rooms, comparisons between them could reveal if there were other factors such as time of day influencing the results, and would allow for them to be accounted for in the analysis.

It is also addressed by taking simple measures of sleepiness from the subjects during each test. If circadian rhythms affecting alertness is an issue, they should show up as systemic differences in sleepiness between the different groups. In that case, their effects could be controlled for in the analysis.

As tutorial times are consistent, the before and after comparisons of the electric lighting and TDDs in VS319 will not be affected.

3.3.1.3 Weather may influence mood, affects the TDDs, and its unpredictability makes scheduling difficult

The variability of the weather causes a number of problems.

Firstly, the weather may impact people's mood (Cohen, 2011). Differences in test responses between different days may thus be influenced by the weather. As the weather is very variable, a mixture of weather conditions can be expected during the study, and getting specific conditions cannot be relied upon.

This is addressed by the control groups. As everyone experiences the same weather, all the groups should be subject to the same effects. Thus, comparisons with the control groups' responses on the same day control for possible effects from this.

Secondly, the weather changes the light levels provided by the TDDs. Overcast days are significantly dimmer than sunny days (section 3.2.1). Ideally, the tests should be run under both overcast and sunny days in order to determine how the varying conditions affect the TDDs, and to allow the results to be more able to be extrapolated, rather than being restricted to specific conditions. However, depending on the weather, such aims may not necessarily be achievable, which would limit the conclusions.

This also means that to have flexibility, more testing opportunities are needed than the actual number of tests planned, so that there is room to cancel a test and run it on another day if the weather is wrong.

3.3.1.4 Differences between groups may affect comparisons between them

Possible demographic differences between the rooms are addressed by surveying the subjects to determine their demographics (section 3.4.1), and then checking and controlling for the potential impacts of factors such as gender on the results.

Another potential issue is that while all the subjects are in the same course (environmental science), those in the test room are from the Architecture stream, while those in the control rooms are from Interior Architecture (VS226) and Building Science (VS322). While they all go to the School of Architecture, and have many courses in common, it is possible that there are systemic differences between the groups and their responses.

Again, this is addressed by comparing all of the rooms under electric lighting, so that any systemic differences between the groups can be identified and accounted for.

A more difficult problem comes from the fact that the different groups may be exposed to certain external stressors on different occasions. Specifically, that the different streams may have assignments from other courses due during the weeks tested. However, avoiding this would require more flexibility and time than is available, given the other constraints on the study. The possible effect of this may be addressed by noting tutorials when the students are stressed by other courses, and carrying out post-hoc assessments of the results.

3.3.1.5 The variability of the tasks the students are doing in their tutorials may make their results less consistent — however it is more representative of real world conditions

A notable difference between this study and similar studies is the variability of subject activity. In typical lighting studies, the subjects carry out a battery of tests that fills the test period (e.g. Boyce et al., 2006a; Knez, 1995; McCloughan et al., 1999). Thus, they are all carrying out the same specific task(s). In this study, however, the students are having their usual tutorial, punctuated only briefly by tests. The students are engaged in a range of computer-based activities, such as building performance simulation, 3D modelling, data analysis, report writing, and talking to their tutors about their assignments. Each tutorial has a different focus and different topics taught. Furthermore, as much of the tutorials involve the tutors assisting the students with their assignments, and as every student is different, the experience of the individual subjects in each tutorial may also be significantly different.

Compared to more normal laboratory studies, this variation in experience may reduce the consistency of the results, especially in measures that look at subject response over the course of the tutorial — such as change in mood.

It is not, however, necessarily a bad thing. This variability of activity is due to it being a real-world situation. As the goal of the study is to get results that can be applied to real world situations, it may in fact be considered an advantage — people carrying out a range of activities that provide different levels of stimulation and experiences is arguably a better representation of many real world situations than people carrying out a specific regimen of tests.

It does of course make analysis more difficult. This is at least partially addressed by running the tests multiple times, and using the control groups to estimate the normal variation in results. If the 'difference' in results under TDDs is within the normal variation of results predicted by the control groups, then we could not be sure that the differences were due to TDDs. The best way to deal with this issue is to run the tests a large number of times to average out any variation. However, the study is limited by the number of available tutorials, as well as the student's willingness to do the same tests again and again.

3.3.1.6 There are only a limited number of days in which tests can be run

Another issue is the limited availability of testing opportunities. The university semester runs for 12 weeks, from the middle of July to halfway through October. There are two tutorials a week. The semester is broken into two 6 week halves, with a 2 week long break in between. As previously discussed, no tests can be run in the weeks when there are assignments due, and it is impractical to run them in the first couple of weeks when everyone's busy getting settled into the course and getting organised into tutorials. This leaves three weeks in the first half of the semester, and four weeks in the second.

The two week long mid-semester break is the most convenient time to "turn-on" the TDDs as it gives flexibility to the workers in terms of scheduling, and because by changing the lighting over the break when most of the students are away from university, we reduce the noticeability of the change.

Thus, there are at most 14 tutorials in which tests can be run, eight of which are in the second half of the semester under TDDs. If each test is aimed to be run twice, once each on sunny and overcast days, then this means that realistically no more than three different tests can be run (depending on how long they are and allowing at least some flexibility for rescheduling due to the weather).

3.3.1.7 Tutorial length limits exposure time

The tutorials limit the subject's exposure to the lighting to short periods of no more than an hour. The tests were run mid-way through the tutorial, after about 25 minutes, as this was the easiest point to integrate them. While this is in line with many lighting studies (e.g. Baron et al., 1992; Knez, 2001; McCloughan et al., 1999; Wang & Boubekri, 2010), it does mean that the lighting has less of an opportunity to influence people and may reduce the likelihood of finding effects (see section 2.4.2.6). It also closes off some avenues of investigation, such as health effects, that need longer term exposures.

3.3.1.8 The specifics of the sample may limit its representativeness

The study's findings may, to some degree, be limited by the population used. The demographics may influence the results (see section 3.4.1), and may affect how much the results can be extrapolated to other groups. The possible influence of the demographics on the results is assessed in the analysis (4.2.2), and is controlled for when necessary.

A basic overview of the demographics of the groups is presented below:

	Proportions of the samples
Female	47-90%
Native English speakers	60-90%
Wears glasses	12-60%
Colour blind	None
Age	All below 30 years old

Table 3-3: Demographics of the samples. As proportions vary between the different rooms and on different days, ranges are given.

The other issue, of whether or not undergraduate student's responses are representative of the broader population, is a common limitation in research (Gifford, 1994), and it can only really be addressed by carrying out more studies on other populations.

3.3.1.9 Getting an adequate sample size

With participation being voluntary, it is impossible to perfectly predict the study's sample size. The question at this stage was if it was *reasonable* to expect a sufficient sample size from a single course, or if more would be needed.

Analysis of precedents suggested that the conditions being compared should each have somewhere between 20-40 subjects (Table 3-4).

Study	Sizes of groups being compared (breakdown in brackets)
Wang & Boubekri (2011)	10 subjects/group
Boubekri et al. (1991)	10-20 subjects/group
McCloughan et al. (1999)	16-32 subjects/group
Knez (1995)	16 -48subjects/group
Knez & Enmarker (1998)	20-40 subjects/group
Knez & Kers (2000)	10 -40subjects/group
Knez (2001)	18-49 subjects/group
Boyce et al. (2006a)	16-90 subjects/group
Newsham et al. (2004)	6-24 subjects/group
Veitch et al. (1991)	~22 subjects/group
Baron et al. (1992)	~11 -45subjects/group

Table 3-4: Sample sizes of lighting studies. Group sizes are given as a range because they are dependent upon how the sample is broken up. A 2x2 study of two light colours and two genders could have the sample divided into two if one just wanted to compare the effects of the light colours, or into four if one was examining the effects of each light colour on each gender separately.

The total number of students in the course was roughly 210. About 120 were in the room with TDDs, while the other 90 were in the two control rooms. Not all of these students regularly attended tutorials. Experience suggested that perhaps 70-80 students in the test room would come to tutorials regularly and be exposed to the TDDs.

Any prediction about participation rates would be speculation. However, a participation rate of 50% would provide a reasonably good sample compared to existing precedents, and did not seem unreasonable.

The actual sample sizes that were achieved are discussed in the analysis where appropriate.

3.4 Control questions

3.4.1 Demographic survey

The demographics were recorded in a survey at the start of the study, before the subjects began the tests.

The demographic characteristics recorded are detailed below, and the survey is reprinted in Appendix E (p138-142).

Characteristic	Reason for including
Gender	Various studies have suggested that men and women *may* respond differently to lighting (Knez & Enmarker, 1998; Knez, 1995, 2001; McCloughan et al., 1999).
Age	It has been suggested that age may affect responses to lighting (Knez & Kers, 2000). This is unlikely to be an issue in this study, but was included for completeness.
Whether or not English is their native language	Performance tests — especially ones involving reading — may be influenced by the subject's fluency with the English language. Psychological research has also suggested that cultural background may affect visual perception (Henrich et al., 2010)
Visual capabilities	Subject's vision is also something that may affect their responses, and is accounted for in a number of studies (e.g. Boyce et al., 2006a; Newsham et al., 2004; Veitch & Gifford, 1996). Subjects are asked whether or not they require glasses and whether or not they are colour blind.

Table 3-5: Demographic characteristics surveyed and their reasons for assessment

3.4.2 Attitudes towards lighting

It is known that an individual's lighting preferences can differ significantly and that this affects their responses to tests (Begemann et al., 1997). It has been suggested that in some cases people may respond positively to different lighting conditions simply because they believe that the lighting *should* be better (Veitch et al., 1991). It has also been suggested that people may respond better to daylighting because they believe that fluorescent lighting is bad for them (Cuttle, 2002).

To check for these kinds of psychological effects, the Lighting Beliefs Questionnaire developed by Veitch and Gifford (1996) was employed. This questionnaire asks subjects 32 questions about how they feel about lighting, such as how important they feel daylight is, and whether or not they think lighting affects their performance.

The questionnaire was delivered after the demographic survey.

3.4.3 Pre-test condition of subjects

A few questions are asked at the start of each test, in order to control for the subject's initial condition.

Questions	Reason for inclusion
Whether or not they have drunk coffee	Drinking coffee may affect subject's alertness and performance (Rautkylä et al., 2010).
Whether or not they have eaten lunch/breakfast	Similarly to coffee, whether or not the subjects have eaten breakfast or lunch before the test could also affect their alertness and performance on tests. If subjects are hungry for instance, they are unlikely to be in a good mood.
Sleepiness	Sleepiness may affect subject performance, or how they feel (see section 3.3.1.2).

Table 3-6: Pre-test questions and their reasons for assessment

3.5 Measuring productivity is problematic

To determine how to measure productivity, we must first ask what productivity is. It is generally considered to be the ratio of what you put *in* to what you get *out* (Heschong Mahone Group, 2003b). In, say, a factory, this is relatively straightforward to measure, as raw material and/or time goes in, and product comes out. In an office setting, however, products are often more nebulous. As Heschong Mahone Group (2003a) discuss, defining the outputs can be quite problematic. For example: how would one define the productivity of a research group? Number of experiments? Words written per day? Such measures are dependent on so many factors that - even if they were reasonable measures of productivity — it would be nigh-impossible to work out, say, whether or not daylighting was improving productivity.

This is not to say, however, that it cannot be done. Depending on the work they do, some organisations may have convenient measures of productivity that can be assessed. A call centre can measure productivity on the basis of how many calls they handle (Heschong Mahone Group, 2003b), retail stores can measure productivity on the basis of sales (Heschong Mahone Group, 1999b, 2003c), while schools may measure it on the basis of student performance (grades) (Heschong et al., 2002; Heschong Mahone Group, 1999b, 2003c). These measures of productivity are, however, the domain of field research that can study the situation over a prolonged period of time. If one is not doing such a study, then these measures may not be able to be used.

Because of these difficulties in directly measuring effects on productivity, it is common in research to use indirect measures instead — testing behavioural or cognitive measures that are believed to be related to productivity, such as absenteeism, memory, or performance on simple tasks (Heschong Mahone Group, 2003b). This has also driven attempts to determine whether or not various commonly studied measures can be linked to productivity — mood being a prime example (e.g. Veitch et al., 2011). Various measures of this kind are discussed later in this chapter.

3.5.1 Productivity cannot be directly assessed in this study

So, can productivity be measured directly in this study?

First, we must ask: what *is* productivity in this instance? As this study is using an educational facility, productivity must be measured by the academic performance of the students — which is measured by their grades, or perhaps, the change in their grades.

Measuring the change in performance here is not possible. To do so would require some kind of standardised test to be administered at the beginning and end of the learning period. This is not part of the courses at the School of Architecture and Design. Grades cannot be compared across different courses. Students can work anywhere — there are four computer rooms in the school, as well as computers in the studios, and many students do their work at home. Trying to control for all of the different environmental variables would be very difficult, requiring an epidemiological study like those of the Heschong Mahone Group (1999b, 2003c). That would need a sample size of thousands (Jackson, 2006a) — something which is not available.

Could the exposure of students to the TDDs be measured and compared to their grades? Again, no. An assessment of how often individuals are exposed to the TDDs could only use vague estimates from the students of how often they use the room, which would have questionable accuracy, and can be expected to vary significantly during the course of the semester, complicating matters further. And, of course, the issue of their limited exposure still applies.

So, to sum up: it is simply too impractical to directly assess productivity in this context.

3.5.2 Therefore *indirect* measures have to be used

So, as productivity cannot be directly assessed in this study, we, like many other researchers, must use indirect measures.

This is not necessarily a bad thing. The trade-off between direct measures of productivity and indirect measures is one of specificity versus generality. The direct measures have the virtue of being highly specific — it is very clear what they mean, and what they apply to. A 20% improvement in student grades is quite straightforward. This is, however, also a limitation — an improvement in student grades may say little about effects on office workers, or retail sales. In contrast, indirect measures such as cognitive performance may suggest effects on a wide range of activities — a 20% improvement in working memory suggests that there are likely to be benefits to anyone doing some task that involves working memory. It does not, however, identify whether or not such an effect has practical impact — it may be likely that it would benefit students, but we cannot tell how much of an effect it would have on their grades. These characteristics should be kept in mind when examining the various tests, and later the results.

3.6 Evaluation and selection of tests

3.6.1 Evaluation Criteria: Time and Likelihood of effect

The methods have been analysed on three criteria:

1) **Time taken** —time is a limited resource so tests that take a significant amount of time (e.g. 20 minutes) are likely to be impractical, or would greatly reduce the number of tests that could be run. 5 minutes is suggested as a reasonable limit that both minimises disruption and is long enough to do tests.
2) **Likelihood of effect** — based on the available precedents, we can estimate how likely it is that the test could find an effect. Relevant precedents are either daylighting studies looking at performance effects, or studies looking at the colour/spectrum of light and its effects.
3) **Value of effect** — as the aim of the study is to find effects of TDDs on people that could provide an argument for people to use them, the potential effects are assessed on their ability to make such a case.

Keys for Likelihood and Value are presented on the right.

Potential size of effect was another possible criterion. However, as large effects are much easier to detect, it is the same as "Likelihood".

Likelihood of effect	
V. Likely	Based on existing knowledge, is almost certain to have an effect.
Positive	There is reason to believe that it will show an effect.
	Has been used in similar studies before, and has often found effects.
Possible	Might have an effect.
	Effects have been found in previous studies, but precedents are too limited to say more. And/or Hasn't really been used in lighting studies, but has been linked to positive mood. And/or Has been used in studies, but overall results have been inconclusive and inconsistent.
Unlikely	It is unlikely that it will have an effect.
	Has been used in lighting studies before, but has not found effects. Additionally, there is reason to believe that this kind of test will not find effects.
Unknown	Not enough relevant precedents to say.

Value of effect	
Good	Effects here can be argued to have economic benefit in terms of productivity, health, or public/employee relations.
Limited	While effects here could be argued to be useful, they are limited to only a few situations.
"Useless"	Effects cannot (easily) be argued to have an economic benefit.

3.6.2 Summary of tests

To determine what tests should be used to examine the effects of TDDs, 28 different tests were investigated.

In the summary table below, the tests have been grouped by first separating out those unlikely to show an effect, and then dividing up the rest based on how long they take.

Detailed summaries and descriptions of the tests are available in Appendix D (p125-137).

Method	Time	Likelihood of showing an effect?	Value of effect?	Notes:
Short (<5min)				
Subjective sleepiness (Rautkylä et al., 2010)	v. short	Positive	Good	Is a single rating scale. Can be done during every test because it's so quick. May be a control.
Mood: questionnaire (Mehrabian, 1974; Watson et al., 1988)	<5 min	Positive	Good	Need to assess at both beginning and end of tutorial to get change in mood. Is linked to many positive effects.
Perception of environment (Veitch et al., 2011)	<5 min	V. Likely	Good	Will almost certainly show results. Is the most basic test of human response to lighting.
Speed of information processing (Lehrl et al., 2007)	~20 sec	Possible	Good	Limited precedents.
Attitude towards university etc. (Russell & Snodgrass, 1991)	~ 1 min	Possible	Good	Possible effect based off mood effects. Likely to be confounded by other factors, and not say anything about lighting.
Digit Span Backwards (Heschong Mahone Group, 2003b)	~ 3 min	Possible	Good	Limited precedents. Assesses attention and short term memory.
Short-term memory: Free recall (Knez, 1995)	<5min	Possible	Good/ limited	Precedents show mixed results. May interact with mood in ways that could limit the value of the results.
Declared seating preferences (Wang & Boubekri, 2010)	v. short	Possible	"Useless"	Can be studied outside of the room/tutorial session. Likelihood of showing an effect most likely linked to changes in light distribution.
Longer (<10min)				
Short-term memory: Article reading (Wang & Boubekri, 2010)	~8 min	Possible	Good	Direct precedents poor.
Motivation/persistence (Boyce et al., 2006a)	<10 min	Possible	Good	May be dependent on positive mood. Precedent is weak.

Creative performance: Alternate uses (Goncalo et al., 2010; Isen et al., 1987)	<10 min	Possible	Good	Possibility based on link to mood. Has not been used in lighting studies before.
Creative performance: Structured imagination (Goncalo et al., 2010; Isen et al., 1987)	~ 7 min	Possible	Good	Possibility based on link to mood. Has not been used in lighting studies before. Difficult to repeat.
Too long				
Colour discrimination: (Boyce et al., 2003)	Hours	V. Likely	Limited	Impractical, expensive, likely results well known. Would say nothing interesting.
Ethics (Banerjee et al., 2012)	~30min	unknown	Limited/ Good	Insufficient precedents. Would create interesting opportunities if effects were found.
Conflict resolution (Baron et al., 1992)	~30min/ ~5min	Positive/ possible	Good	Time taken may be shortened if fewer scenarios are used. However this may make the results less reliable.
Problem solving: embedded figure task (Knez, 1995)	~35min	Possible	Good	No daylighting precedents.
Long-term recall (Knez, 1995)	55 min + gap	Possible	Good	No daylighting precedents.
Creative performance: Torrence tests (Dow, 2003; Isen et al., 1987)	~30min	Possible	Good	Difficult to use, would cost. Possibility based on link to mood.
Observed seating preferences (Wang & Boubekri, 2010)	Long	Possible.	"Useless"	Can be studied outside of the room/tutorial session. Likelihood of having an effect most likely linked to changes in light distribution. Would be confounded with many factors.
Unlikely to get any results				
Performance: arithmetic/underlining (Boray et al., 1989; Veitch et al., 1991)	~2 min	Unlikely	Good	Precedents are limited and are negative. No daylighting precedents.
Absenteeism (Heschong Mahone Group, 2003a)	NA	Unlikely	Good	Even if there was a potential effect, it would likely be masked by other confounding variables.
Performance: simple clerical (Boyce et al., 2006a)	~15min	Unlikely	Good	Has had little success in lighting studies. It is argued that such tests are unlikely to find effects.
Performance: complex cognitive (Boyce et al., 2006a)	~40min	Unlikely	Good	Has had little success in lighting studies. It is argued that such tests are unlikely to find effects.
Visual performance: Landolt rings (Boyce et al., 2003)	~5min	Unlikely	Limited	Unlikely to show any effect unless looking at near-threshold tasks. Failing that, any effect would be due to illumination differences.

Health (Mills et al., 2007)	Variable	Unlikely	Good	Need longer term exposure — people working all day, for weeks.
Eye discomfort (Newsham et al., 2004)	~1 min	Unlikely	Good	Need longer term exposure — people working all day, for weeks.
Stress (Fostervold & Nersveen, 2008)	short	Unlikely	Good	Need longer term exposure — people working all day, for weeks.

Table 3-7: Comparative summary of possible methods

3.6.3 Discussion: Test selection

3.6.3.1 Most of the tests could find valuable effects

Most of the effects found by the tests could be used to argue that TDDs provide valuable benefits. Therefore, the discussion of the test selection will focus on the aspects of 'Time' and 'Likelihood of finding effects'.

3.6.3.2 There are good matches between 'Time taken' and 'Likelihood of effect'

Fortunately, the 'Time taken' and 'Likelihood' groups match up well, with a number of the shorter tests having better chances of finding effects, and most of the longer tests being less likely to find effects. This makes it easier to select the most appropriate tests.

3.6.3.3 'Perception', 'Mood', and 'Sleepiness' are obvious choices

Surveying people on their **perception** of the lighting and room is an obvious test to run. It is one of the standard tests used in lighting studies (e.g. Boyce et al., 2006a; Knez, 1995; Veitch et al., 2011), as it is the only way to find out what people think about the lighting, and whether or not they perceive it to be different. It is very useful for identifying details about how people are perceiving the lighting, as well as potential problems such as glare. It is the test most likely to show effects, and it can be carried out in a short period of time. Positive results could suggest benefits to the image of the building, and possible effects on job satisfaction and productivity (Veitch et al., 2011).

Mood surveys are another obvious choice in this study. They are also a 'standard' test in lighting studies (e.g. Knez & Enmarker, 1998; McCloughan et al., 1999; Plitnick et al., 2010; Viola et al., 2008), as positive mood has been linked to many desirable outcomes , including improved productivity (Harter et al., 2002; Veitch et al., 2011) and pro-social behaviour such as greater generosity and more cooperative conflict resolution (Russell & Snodgrass, 1991). Many of the tests examined above are based on effects that may be related to positive mood. Thus, positive effects of mood could be linked with, and could reinforce, positive effects in the other tests. There are a number of successful precedents suggesting a reasonable likelihood of the test finding effects (Boyce et al., 2003), and the tests are relatively short — though they are complicated by needing to be carried out at both the beginning and end of the tutorial session. It should be noted that precedent also suggests that it is quite possible that it will not find any effects: mood surveys have had a mixed success rate, as was discussed in the literature review.

Lastly, **sleepiness** is a test that takes almost no time at all, consisting as it does of a single question ("rate how sleepy you feel on this scale"). As it has reasonable odds of finding an effect, and is being used as a control anyway (see Section 3.4.3), it is an obvious choice. It may also corroborate the measurements of arousal in mood, and can be used to see if any trends emerge over multiple tutorials.

These three tests are the most obvious choices, and are all tests that are commonly used in lighting studies. There is, however, one last category of tests that are commonly used in lighting studies — the **performance** tests.

3.6.3.4 Performance tests have had limited success in these kinds of study

The problem with the cognitive performance tests (other than the fact that many of them take a significant amount of time) is that lighting studies have generally had only limited success with them, and because there are so many *different* performance tests, there is limited precedent for any specific ones. A number of them (mainly ones looking at simple task performance) have repeatedly failed to find effects in lighting studies (Boyce et al., 2006a; Newsham et al., 2004; Veitch et al., 2008), and so are unlikely to find effects here. It has been suggested by Boyce et al. (2006a) that in a testing situation it is relatively trivial for people to perform at their maximum potential for the short period they are there for — and so as long as the light is decent, it will have little impact on their performance. In contrast, in the real world, people may slack off, or lose concentration over longer periods, and lighting may influence these behaviours more (Boyce et al., 2006b).

Other tests (mainly memory based ones) have had more success. However, their results have not been consistent, and when effects are found they tend to be subtle —perhaps for the reasons discussed above (e.g. Knez & Enmarker, 1998; Knez & Hygge, 2002; Knez, 1995, 2001).

Lastly, there is a very limited selection of performance tests under daylighting, so most of the precedents are of studies using electric lighting and looking at the effects of light colour on performance.

3.6.3.5 'Information processing' and 'Digit Span Backwards' are the best options for performance tests

Despite the performance tests' generally poor record, two tests have been selected for use in this study. This has been done for several reasons:

1) Because it's standard practice. As discussed in the literature review, the different human factors tested for lighting effects can be divided into four categories: Perception, mood, health, and performance. Health needs longer-term field studies, so laboratory lighting studies tend to measure perception, mood, and performance (e.g. Boyce et al., 2006a; Knez, 1995; Newsham et al., 2004).
2) Because it provides more variation in the tests. As discussed earlier (section 3.3.1.1), we want to provide a variety of tests that can be mixed up to keep the students interested and participating in the study. Just 'mood' and 'perception' isn't really enough to do this.
3) And, because performance is what we're really interested in. Performance tests are, at least conceptually, the closest we can come to productivity tests in this context — and indeed the requirements that we have for lighting in our buildings are based around enabling good performance (e.g. NZS 1680.1).

For these reasons, it is felt to be important to test performance.

The first test chosen is **Digit Span Backwards**. It is a widely used test that measures working memory and people's ability to concentrate (Heschong Mahone Group, 2003b). It is one of the shortest performance tests, and of the memory tests, is considered to be the most promising. In a field study in an office building, the Heschong Mahone Group (2003a) used it to find a significant — and substantial — effect of daylight on performance.

Secondly: **Speed of information processing**. This test can also be carried out fairly quickly, and has the virtue of having been used to find effects of light on performance in a short study with short exposure times (Lehrl et al., 2007) — suggesting that it may work well in the limited timeframes available in this study.

These tests are both very short — short enough, in fact, that they could be run together in one five minute test block, which is very convenient for this study.

Using multiple tests provides a greater chance of finding effects, and if they both find effects then they would support each other, and strengthen the conclusions.

The issue with both these tests is that they have very limited precedents — specifically, one each. However, the tests that have been used more often in lighting studies have tended to take longer, and have tended not to find effects. So it was decided that these two tests were more appropriate.

3.6.3.6 Other tests take too long, or are unlikely to find effects, or are simply much less promising

The remaining tests are generally more time consuming than the selected ones, and are less promising, with much weaker precedent for successful results. Many are based on links to possible effects of positive mood, and have not actually been assessed in lighting studies themselves.

Roughly half the tests are ruled out for either requiring too much time to be practical (30 minutes or more) or simply being unlikely to find any effects.

3.7 Details of the tests chosen for this study

The selected tests are briefly described and discussed below. The actual tests and survey forms may be found in Appendix E.

3.7.1 Perception of environment

People's perception of the lighting and environment is assessed using surveys asking them to rate how they feel about it. Two surveys are used here, both used previously in studies by the National Research Council Canada (e.g. Veitch & Newsham, 2000; Veitch et al., 2011). By using the same tests, we make it easier to compare results across the different lighting studies.

One test is on lighting quality, and the other on room attractiveness.

3.7.1.1 Lighting quality

The lighting quality survey consists of seven questions to which subjects rate their response on five-point Likert scales. Five of the questions deal with perceived lighting quality and satisfaction, and two with glare problems (e.g. "how satisfied are you with the lighting in the room?"). Scores across the questions are averaged to give the overall lighting quality and glare ratings.

An eighth question was added to the survey asking the subjects to rate the daylight available to them. This was asked because Carter & Marwaee (2009) suggested that people may not be appreciating the light through TDDs as daylight — so the question was added to find out if they are recognising it as daylight, and if they think it is good quality.

3.7.1.2 Room appearance

Room appearance is assessed using semantic differential scales. The survey is based on tests first run by Flynn (1977), and has been refined down to eight paired descriptors with high internal consistency and reliability by Dr. Veitch of the National Research Council Canada. Subjects are given pairs of words (e.g. attractive – unattractive) on either end of a nine-point scale. They are then asked to describe how they feel about the room by placing a mark on the scale. The more that one adjective describes their feelings, the closer they put their mark to it. Five of the pairs describe attractiveness and likeability (attractive/unattractive, cheerful/sombre, pleasant/unpleasant, beautiful/ugly, like/dislike), while the other three describe illumination (radiant/gloomy, distinct/vague, bright/dim). Ratings are averaged to provide overall scores for each of the two factors.

3.7.2 Cognitive performance

3.7.2.1 Digit Span Backwards

In Digit Span Backwards, subjects are given a random string of numbers, one after another, with each number presented one second apart. They then have to repeat the numbers backwards. They are first given two strings of three numbers, and then a set of four numbers, and so forth, progressing up to nine numbers. They are scored on the longest set of strings that they can accurately remember, with half a point given if they can only succeed at one string in the set (Heschong Mahone Group, 2003b). The test ends when they fail on both of the strings in a pair.

The strings were randomly generated, and then edited to make sure there were no easy-to-remember patterns or repetition such as "1-2-3" or "5-5-5".

The same strings of numbers were used in both the 'before' and 'after' test sessions. This was not felt to be a problem, as there was sufficient time between tests (several weeks) that it was felt to be unlikely that the students would remember the strings, and it was unlikely that they would be interested in taking the effort required to cheat. Moreover, keeping them the same avoids having to deal with the possibility that some strings may be harder to remember than other strings, which would interfere with the results.

3.7.2.2 Processing Speed

In the processing speed test, subjects are shown a line of 25 random different letters. The sequences contained every letter in the alphabet apart from "W", as testing suggested that its long pronunciation could be a stumbling block when reading through the letters. The letters are light grey on a dark background. They are instructed to read them, from left to right, as fast as possible. They are given 4 seconds to read, and are then asked to write down the last two letters they read (Lehrl et al., 2007). This measures how fast they can process "bits" of information, and has been associated with fluid intelligence and IQ (Lehrl et al., 2007).

The subjects repeated the test 10 times in a session, and their score was averaged, as testing suggested that this would provide more stable and reliable results. It also allowed subjects to make a mistake once or twice, and still have a viable overall performance score.

Subject's results were eliminated if more than two of their test scores had to be eliminated. Test scores could be eliminated for two reasons:

1) If they failed to input a valid pair of consecutive letters — which meant that their score could not be accurately assessed.
2) If they achieved the maximum score. Pilot testing suggested that legitimately achieving the maximum score was improbable — and discussions raised the possibility that people might "cheat", and just read the last two letters on the line. Thus, it is assumed that anyone getting the maximum score had not correctly carried out the test, and their score was not counted.

Like digit span backwards, the same sequences of letters were used in both tests, for the same reasons.

3.7.3 Mood: the Russell and Mehrabian semantic differential scale

Mood is assessed using the Russell and Mehrabian semantic differential scale (Russell & Mehrabian, 1977). This was selected because it was used in the studies by the National Research Council Canada (e.g. Boyce et al., 2006a; Veitch & Newsham, 2000; Veitch et al., 2011), and because it has been argued that the main alternative (PANAS (Watson et al., 1988)) does not work very well in these studies (Knez, 2001).

The mood scales ask subjects to rate how they feel, on scales running between semantic pairs — for example: happy – unhappy. Two mood "factors" were assessed: "pleasure", and "arousal". The third factor of the scale (dominance) was not assessed, because of time constraints, and because it was not felt to be very important, as the subjects had no control over the lighting in either test situation, and so it would be unlikely to vary.

Each factor is assessed using six semantic pairs, and the average score on them gives the overall rating of pleasure or arousal. Pleasure is assessed by how much a person feels happy, satisfied, hopeful, contented, pleased, and relaxed. Arousal is assessed by how much they feel stimulated, excited, frenzied, jittery, awake, and aroused.

Mood is assessed twice in a tutorial — once at the beginning, and once at the end. This allows the measurement of the *change* in mood of the subjects, which helps to reduce the confounding effects of what their mood was like before coming to tutorial (Boyce et al., 2003). This means that for the mood tests, the first test is run earlier than the other tests, at 10-15 minutes in. The interval between tests is roughly 35 minutes, which is shorter than usual (e.g. 80 minutes (Knez, 2001), 45 minutes (N. Wang & Boubekri, 2011)), and could reduce reliability of the change in mood. However, such a short period does have some precedent (e.g. McCloughan et al., 1999).

3.7.4 Sleepiness

Sleepiness is assessed using the Karolinska Sleepiness Scale (Akerstedt & Gillberg, 1990). Subjects are asked to rate how sleepy they feel on a nine-point scale from "very alert" to "very sleepy". As it is so short, it is carried out in every test session.

4 RESULTS: ARE TUBULAR DAYLIGHTING DEVICES BETTER?

To answer the question of whether TDDs have better effects than electric lighting, there are several issues that need to be addressed:

1) Are people's responses different under TDDs?
2) Can we be confident that the differences are caused by the lighting?
3) What are the mechanisms involved? Could the effects be replicated with different electric lighting?
4) Is the electric lighting looked at in this study typical for offices and educational buildings?
5) Following on from the above, what can now be said about TDDs? Is it a superior option? In what ways?

4.1 Are people's responses different under TDDs?

This section presents the results of the tests. Specifically:

1) People's perceptions of the room and lighting
2) Their performance on the cognitive performance tests
3) The assessments of their mood
4) Their reported alertness

4.1.1 Perception of the room and lighting

The surveys of people's perceptions of the rooms and lighting provide:

1) A rating of the lighting quality — how satisfied they are with the lighting in their workspace.
2) A rating of the level of glare.
3) A rating of the quality of the daylight available to them.
4) A rating of the appearance of the room described by two factors: attractiveness and illumination.

The results for each rating are presented separately, and then the overall findings are discussed.

The surveys were done three times — once in the first half of the semester, with room VS319 using electric lighting, and twice in the second half with VS319 under TDDs. One round under TDDs was on a sunny day, the other on an overcast day (Figure 4-1). Due to tutorial complications one of the control groups, VS322, missed the first test round, and so only had two rounds of surveys.

	1st half	2nd half	
	Test round 1	**Test round 2**	**Test round 3**
VS319	Electric lighting	TDDs – sunny weather	TDDs – overcast weather
VS322		Electric lighting	Electric lighting
VS226	Electric lighting	Electric lighting	Electric lighting

Figure 4-1: Diagram outlining the rounds of perception surveys

To provide a simple overall comparison of the rooms, each room's results from the different rounds was combined, with individuals' responses being averaged across the rounds (Table 4-1). Averaging the subjects' responses across multiple tests improves their reliability, reducing their natural variation. Combining the rounds also maximised the sample sizes, thus providing greater statistical power.

VS319 (TDDs)	Round 2 + Round 3
VS319 (Electric)	Round 1
VS226	Rounds 1 + 2 + 3
VS322	Rounds 2 + 3

Table 4-1: Survey rounds combined to provide overall scores for the different rooms/conditions

Paired comparisons were made by matching up individual subject's responses in different rounds and measuring their change in response. This measure controls for the differences between people — making it more sensitive. Looking at the data in these two different ways provides more rigorous analysis, as they each help shore up the others' shortcomings: for the simple averages, the effects of individual variation; for the paired comparisons, a smaller sample size and the fact that they cannot compare different rooms.

Statistical significance of the differences between conditions was assessed using the Student's T-test (two-tailed)(D. S. Moore, 2000).

4.1.1.1 Lighting quality

Below, in Figure 4-2, the overall mean ratings of lighting quality are shown for each of the four conditions. Next to it, in Figure 4-3, a statistical summary shows the significance of the differences between the different rooms/conditions.

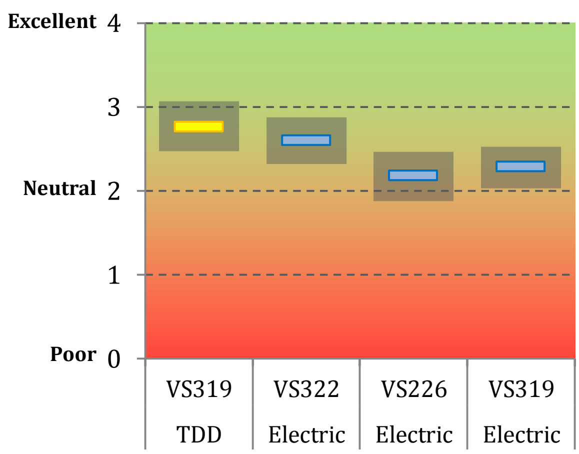


Figure 4-2: Mean ratings of lighting quality for the different situations. Shaded areas show the margin of error (95% confidence). The background gradient provides a visual guide: green is good, and red is bad.

		TDD	ELECTRIC		
		VS319	VS322	VS226	VS319
		n 29 m 2.77 sd 0.67	*n* 23 m 2.60 sd 0.56	*n* 26 m 2.17 sd 0.63	*n* 49 m 2.28 sd 0.76
TDD	VS319		dif. 0.17 t 1.02 p ns	dif. 0.60 t 3.42 p <0.005	dif. 0.49 t 2.98 p <0.005
ELECTRIC	VS322			dif. 0.43 t 2.51 p <0.02	dif. 0.32 t 2.00 p <0.05
	VS226				dif. -0.11 t -0.65 p ns
	VS319				

Figure 4-3: Significance of the differences between the ratings of lighting quality in the different situations and associated statistics. dif. = difference between the means, t = t-value of the difference, p = p-value. Note: significant differences are highlighted green, while non-significant ones are grey.

These comparisons show several things:

1) The TDDs *significantly* improved the lighting quality in VS319.
2) The electric lighting in both VS319 and VS226 is of similar quality.
3) VS322's electric lighting has *significantly better* quality than that in VS226 and VS319. Its lighting quality is comparable to the TDDs.

A paired comparison was run assessing the change in individuals' responses from electric lighting to TDDs (Table 4-2, below).

	n	Mean difference	SD	t(24)	significance
VS319: TDDs vs. Electric	24	0.70	0.99	3.54	p<0.005

Table 4-2: TDDs vs. Electric lighting in VS319: Paired comparison of lighting quality using the differences in individuals' ratings in each situation. Columns are, from left to right: sample size, mean change in rating from electric lighting to TDD, standard deviation of the change in ratings, t-value, and p-value.

The paired comparison confirms that the lighting quality was significantly improved under TDDs, with a significant *positive* change in rating.

To check whether the change could be because of natural variation or possible external factors, the results are broken down to see how they varied across the different rounds (Figure 4-4 below). Unfortunately, due to VS322 missing the first round, and having very poor participation in the third ($n = 4$), only VS226 can be used to look at the natural variation in an unchanging room — resulting in a relatively small sample size (sample sizes by round: 17, 14, 11). As can be seen though, the average ratings are reasonably consistent across the rounds, giving no reason to believe that the improvement in lighting quality under TDDs is *not* due to the change in lighting.

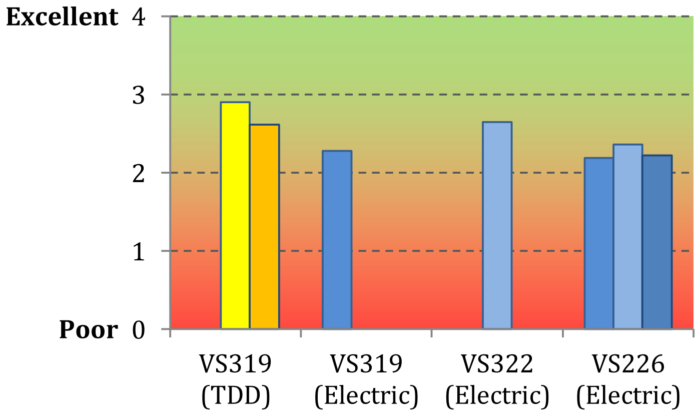


Figure 4-4: Lighting Quality results broken down by test round for each room.

The last comparison is between the sunny and overcast days for the TDDs:

	n	Mean difference	SD	t	significance
Sunny vs. Overcast	15	0.09	0.84	0.42	ns

Table 4-3: Sunny vs. Overcast day for TDDs: Paired comparison of lighting quality using the differences in individuals' ratings of lighting quality in each situation.

Paired comparison of the two days finds no significant difference in lighting quality (Table 4-3). If, however, the sunny and overcast days are each *separately* compared to the electric lighting (Table 4-4), then we see that the improvement in lighting quality is lower on the overcast day, having only marginal (p<0.1) significance — suggesting that lighting quality *might* be slightly better on sunny days than overcast days for TDDs. Any difference would be subtle however, and it could very well just be due to natural variation.

	n	Mean difference	SD	t	significance
Sunny day vs. electric	19	0.79	0.98	3.52	p<0.005
Overcast vs. electric	18	0.5	1.12	1.90	p<0.1

Table 4-4: Paired comparisons of lighting quality between TDDs and electric lighting separated into the sunny and overcast days.

Overall, these results suggest that, while adding TDDs to VS319 did improve matters, the results are more demonstrative of the poor quality of the electric lighting in VS226 and VS319. As VS322 shows, electric lighting with parabolic reflectors and louvers can have similarly good lighting quality (Figure 3-2,Figure 3-3).

The lack of any significant difference between sunny and overcast days suggests (somewhat surprisingly) that light level has little effect on lighting quality. The fact that the lighting was just as good when overcast is interesting, considering that the light levels then were significantly lower than those recommended by NZS 1680.1 (~190 vs. 320 lux).

4.1.1.2 Glare problems

Overall, there is no significant difference in glare levels in the different conditions (Figure 4-5, Figure 4-6). The paired comparison confirms that the TDDs do not have a significant impact on the ratings of glare. This *was* a potential concern, as when the TDDs were installed on a sunny day, the IT department in the School of Architecture expressed concern that the apparently higher light levels might result in glare problems. As discussed in section 3.2.1, while the light levels on the desktops were similar to those under the electric light, the vertical illuminances on the screens and walls were higher, making the room appear much brighter.

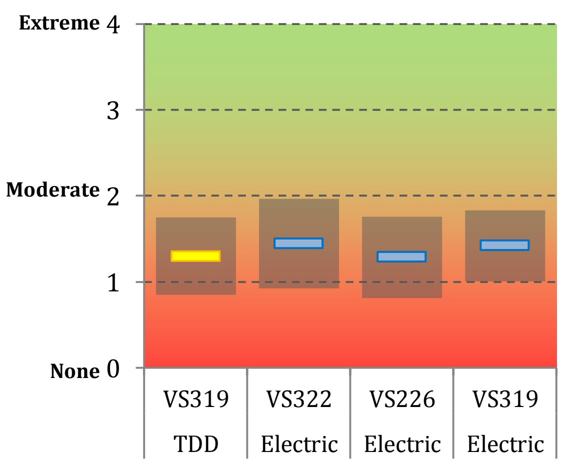


Figure 4-5: Mean ratings of glare for the different situations.

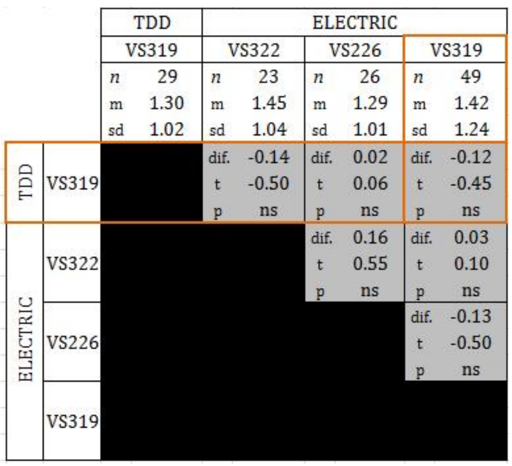


Figure 4-6: Significance of the differences between the ratings of glare in the different situations and associated statistics.

	n	Mean difference	SD	t	significance
VS319: TDDs vs. Electric	24	-0.13	0.92	-0.71	ns

Table 4-5: TDDs vs. electric lighting in VS319: Paired comparison of glare using the differences in individuals' ratings in each situation.

Breaking down the results by round, however, reveals a possible round effect, as both the groups in both VS319 and VS226 had lower ratings in the third round (Figure 4-7).

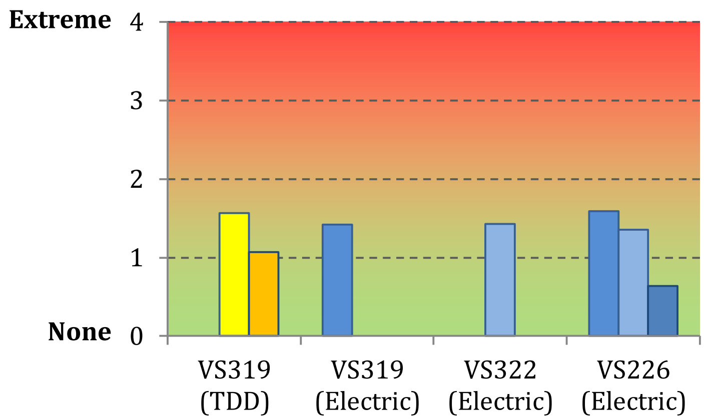


Figure 4-7: Glare results broken down by test round for each room.

Closer analysis shows several things:

1) In VS319:
   a. Paired comparison shows a significant difference in glare ratings between the second and third rounds (sunny and overcast) (Table 4-6).
   b. The third round is not, however, significantly different from the first round (electric) (Table 4-6).
   c. The difference between the average ratings on the second and third rounds *also* is not significant (dif. = -0.5, t = 1.5, ns).
2) In VS226:
   a. Average ratings of glare are significantly lower in the third round (dif. = -0.85, t(40) = 2.9, p<0.01).
   b. Paired comparison between the third round and the previous ones does *not* show a significant difference (dif. = -0.84, t(7) = 1.77, ns), but that might just be because of its very low sample size (*n* = 8).

Overall, the data raises the *possibility* that glare problems from TDDs may be slightly lower on overcast days — which makes sense given the lower light levels — but we cannot be confident about it, as other comparisons do not show significant differences, and VS226 suggests a possible round effect.

	n	Mean difference	SD	t	significance
Sunny vs. overcast day	15	0.59	0.93	2.54	p<0.05
Sunny day vs. electric	19	0	0.67	0	ns
Overcast vs. electric	18	-0.3	1.18	1.10	ns

Table 4-6: Sunny vs. Overcast days for TDDs: Paired comparisons of glare between each other, and against electric lighting separately.

Ultimately, the levels of glare are similar across the rooms, and TDDs are not significantly different to the electric lighting, either on overcast or sunny days. Moreover, glare in all the rooms is generally low, so effects here are likely to be of limited benefit.

4.1.1.3 Daylight quality

The ratings of daylight quality are about what would be expected: very low for the non-daylit rooms; significantly higher under the TDDs (Figure 4-8, confirmed by paired comparison (Table 4-7)). There are no significant differences between the three non-daylit rooms. This shows that the students were aware that the TDDs were providing daylight.

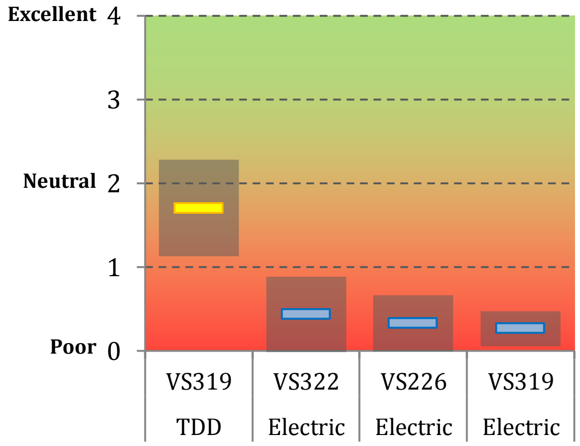
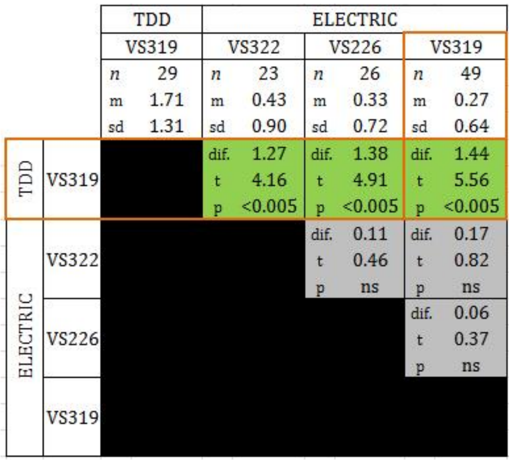


Figure 4-8: Mean ratings of daylight quality for the different situations.



		TDD VS319	ELECTRIC VS322	ELECTRIC VS226	ELECTRIC VS319
		n 29 m 1.71 sd 1.31	*n* 23 m 0.43 sd 0.90	*n* 26 m 0.33 sd 0.72	*n* 49 m 0.27 sd 0.64
TDD	VS319		dif. 1.27 t 4.16 p <0.005	dif. 1.38 t 4.91 p <0.005	dif. 1.44 t 5.56 p <0.005
ELECTRIC	VS322			dif. 0.11 t 0.46 p ns	dif. 0.17 t 0.82 p ns
ELECTRIC	VS226				dif. 0.06 t 0.37 p ns
ELECTRIC	VS319				

Figure 4-9: Significance of the differences between the ratings of daylight quality in the different situations and associated statistics.

	Mean difference	SD	t(24)	significance
VS319: TDDs vs. Electric	1.34	1.41	4.74	p<0.005

Table 4-7: TDDs vs. electric lighting in VS319: Paired comparison of daylight ratings using the differences in individuals' ratings in each situation.

Breaking down the results by round shows nothing of note — the overall results are very clear and straightforward (Figure 4-10).

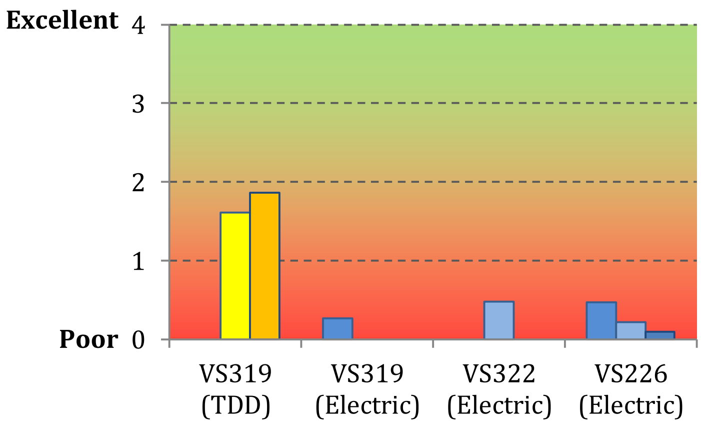


Figure 4-10: Daylight quality results broken down by test round for each room.

There is no significant difference between sunny and overcast days (Table 4-8). This demonstrates that the perception of the quality of the daylight was not being affected by light level - and it suggests that good daylighting does not necessarily require lots of daylight.

	n	Mean difference	SD	t	significance
Sunny vs. overcast day	15	-0.06	1.29	0.19	ns
Sunny day vs. electric	19	1.05	1.47	3.12	p<0.01
Overcast vs. electric	18	1.56	1.46	4.51	p<0.005

Table 4-8: Sunny vs. Overcast days for TDDs: Paired comparisons of daylight ratings between each other, and against electric lighting separately.

There are also a couple of other interesting points:

1) Some of the students gave non-zero daylight ratings in the non-daylit rooms. This is possibly a way of saying that the lack of daylight is acceptable to them, but this speculation would require further study to confirm.
2) The average rating of the daylight quality under TDDs is still not that good, being slightly less than “neutral”. This suggests that the students may expect more from their daylight than the TDDs are providing. Further examination of this issue would require a more in-depth study than was possible here, examining more variations in daylighting.

4.1.1.4 Room attractiveness

The results show that the TDDs make VS319 appear significantly more attractive — shown both in the comparisons of the averages (Figure 4-11) and in the paired comparison (Table 4-9). Note also that none of the rooms are particularly attractive. With electric lighting, they can be characterised as “somewhat unattractive” and the TDDs, while better, only raise that to “so-so”, or neither attractive nor unattractive (Figure 4-11).

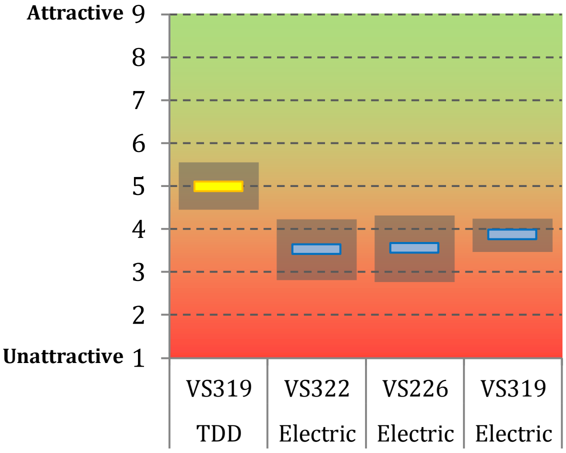


Figure 4-11: Mean ratings of the room attractiveness for the different situations.

		TDD	ELECTRIC		
		VS319	VS322	VS226	VS319
		n 29 m 5.00 sd 1.25	*n* 23 m 3.51 sd 1.41	*n* 26 m 3.54 sd 1.67	*n* 49 m 3.85 sd 1.18
TDD	VS319		dif. 1.48 t 3.97 p <0.005	dif. 1.45 t 3.63 p <0.005	dif. 1.15 t 4.01 p <0.005
ELECTRIC	VS322			dif. -0.03 t -0.07 p ns	dif. -0.34 t -0.99 p ns
	VS226				dif. -0.31 t -0.83 p ns
	VS319				

Figure 4-12: Significance of the differences between the ratings of room attractiveness in the different situations and associated statistics.

	n	Mean difference	SD	t	significance
VS319: TDDs vs. Electric	24	1.34	1.95	3.44	p<0.005

Table 4-9: TDDs vs. electric lighting in VS319: Paired comparison of room attractiveness using the differences in individuals’ ratings in each situation.

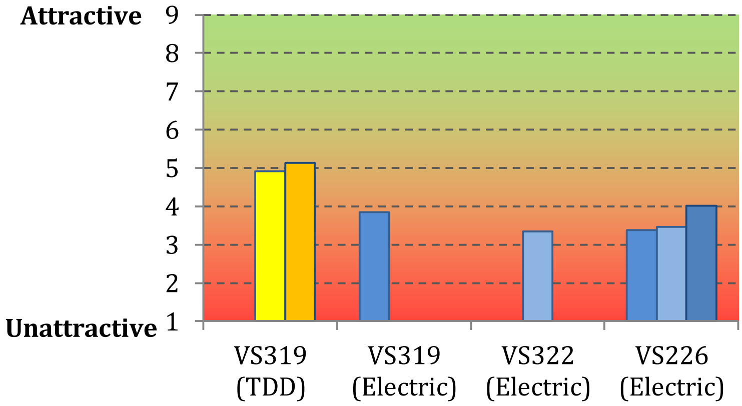


Figure 4-13: Room attractiveness results broken down by test round for each room.

Breaking down the results by round shows that the ratings of room attractiveness are fairly consistent (Figure 4-13).

	n	Mean difference	SD	t	significance
Sunny vs. overcast day	15	0.11	1.23	0.37	ns
Sunny day vs. electric	19	1.16	1.79	2.82	p<0.02
Overcast vs. electric	19	1.01	1.97	2.17	p<0.05

Table 4-10: Sunny vs. Overcast day for TDDs: Paired comparisons of attractiveness ratings between each other, and against electric lighting separately.

Paired comparison shows that there is no significant difference in ratings of attractiveness between sunny and overcast days for TDDs (Table 4-10), suggesting that light levels, at least within the range looked at here, do not significantly affect the attractiveness of a space.

4.1.1.5 Perceived illumination

The results show that the TDDs made the room appear significantly brighter than with electric lighting, with the TDDs being perceived as significantly brighter than any of the rooms with electric lighting (Figure 4-14,Figure 4-15, further supported by paired comparison (Table 4-11)).

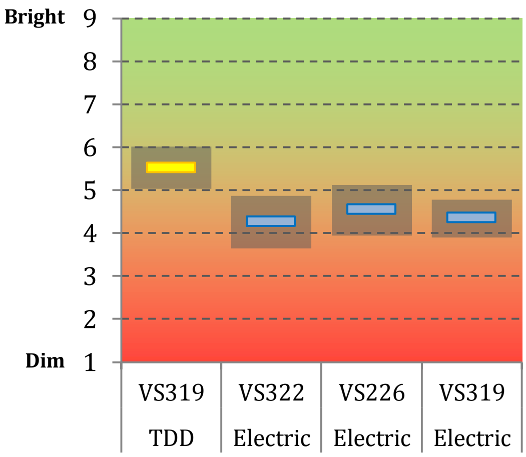


Figure 4-14: Mean ratings of the illumination for the different situations.

		TDD	ELECTRIC		
		VS319	VS322	VS226	VS319
		n 29 m 5.52 sd 1.12	*n* 23 m 4.26 sd 1.21	*n* 26 m 4.54 sd 1.26	*n* 49 m 4.34 sd 1.33
TDD	VS319		dif. 1.26 t 3.85 p <0.005	dif. 0.99 t 3.06 p <0.005	dif. 1.18 t 4.20 p <0.005
ELECTRIC	VS322			dif. -0.28 t -0.78 p ns	dif. -0.08 t -0.25 p ns
	VS226				dif. 0.20 t 0.63 p ns
	VS319				

Figure 4-15: Significance of the differences between the ratings of illumination in the different situations and associated statistics.

	n	Mean difference	SD	t	significance
VS319: TDDs vs. Electric	24	1.10	1.48	3.71	p<0.005

Table 4-11: TDDs vs. electric lighting in VS319: Paired comparison of illumination ratings using the differences in individuals' ratings in each situation.

Breaking down the results by round shows that the ratings of illumination are fairly consistent (Figure 4-13). There is a suggestion that illumination might be slightly lower on overcast days, but it is still significantly higher than it is under electric lighting (dif. = 0.90, $t(69) = 2.9$, $p<0.01$).

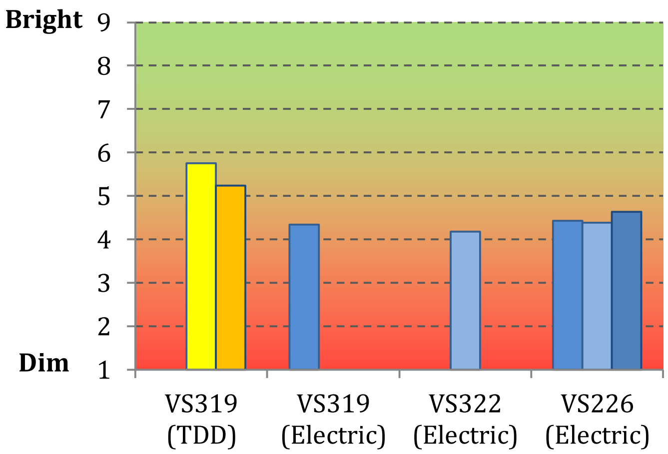


Figure 4-16: Illumination results broken down by test round for each room.

	n	Mean difference	SD	t	significance
Sunny vs. overcast day	15	0.65	1.01	2.55	p<0.05
Sunny day vs. electric	19	0.98	1.22	3.50	p<0.005
Overcast vs. electric	19	0.56	1.64	1.44	ns

Table 4-12: Sunny vs. Overcast day for TDDs: Paired comparisons of illumination ratings between each other, and against electric lighting separately.

Paired comparison of the sunny and overcast days shows that the room on the sunny day is significantly brighter (Table 4-12) — as would be expected from the higher light levels. Paired comparison between the TDDs on the overcast day and electric lighting does *not* confirm the difference as significant, suggesting that the difference between TDDs and electric lighting on overcast days may be more subtle, or that it might not be reliable.

4.1.1.6 Discussion

The key findings of the perception surveys are as follows:

1) TDDs significantly increased the lighting quality, attractiveness, and perceived brightness of the room in VS319. All three of the rooms with electric lighting were significantly less attractive and appeared dimmer.
2) The electric lighting in VS322, however, provided similar lighting quality to the TDDs.
3) Subjects recognised the TDDs as providing daylight, but it was not necessarily considered great quality.
4) TDDs did not create increased glare, despite higher vertical light levels on sunny days.

5) With electric lighting, VS319 generally has a similar environment to the two control rooms, with similar ratings of attractiveness, daylight quality, and glare. However, VS322 has better lighting quality than the electric lighting in VS226 and VS319.
6) TDDs provide the same quality of light, and make the room just as attractive, on both sunny and overcast days. However, the light is dimmer on overcast days, and on such occasions TDDs might not necessarily be brighter than electric lighting.

From this, two key conclusions can be drawn:

1) The contrast between the lighting quality ratings and the room attractiveness ratings is interesting. The fact that VS322 has higher lighting quality, but is no more attractive than the other electrically lit rooms, indicates that they are being influenced by different factors. It shows that lighting that is satisfactory for people’s work does not in itself provide a pleasant environment. It may be regarded as a strength of the TDDs that they manage to do both here.
2) TDDs are generally the best option here, having lighting quality no worse than the best electric lighting in this study, and making the room significantly more attractive. They work well even on overcast days, showing that they can be consistently effective regardless of the weather conditions.

4.1.2 Cognitive performance

Cognitive performance was assessed by two tests, of working memory (the Digit Span Backwards test) and processing speed. One round of testing was carried out in the first half of the semester (when VS319 was under electric light), followed by a second round in the second half of the semester, when VS319 was under TDDs.

As discussed in the methodology (3.7.2), some subjects' results were removed as they failed to correctly carry out the tasks. In the Processing Speed test, 18 subjects had their results removed, leaving a total sample of 153.

Three subjects' working memory scores from VS226 in the second round were removed because the subjects appeared to have not bothered to do the test — they had given clearly incorrect answers (e.g. "111") that allowed them to end the test as fast as possible with a score of 0. However, the first round had showed that they had the ability to perform the task at a significantly higher level.

The basic analysis is the same as that carried out on the perception results. However, the identification of a possible confounding effect of *round* on working memory performance necessitated additional statistical analysis of the results. A linear regression model was used to further analyse the possible effects of the round, the different rooms, and the TDDs. It allowed the effects of the different factors to be assessed simultaneously, and thus to determine how much of an impact each had on the variance in the results whilst controlling for the others.

4.1.2.1 Working memory (Digit Span Backwards)

Looking at the overall results (Figure 4-17), the key points of note are:

1) Under the TDDs, the scores were the lowest (Figure 4-17). They were *significantly* lower than those in VS226, and *marginally* significantly lower (p<0.1) than they had been previously in the room under electric lighting (Figure 4-18). However, they were *not* significantly lower than in the other control room (VS322).
2) There were no significant differences between the rooms with electric lighting.

This suggests that TDDs may have *reduced* performance on the working memory test — though the marginal significance of the difference in VS319, and the lack of significant difference to the second control group suggest that it may not be reliable, or that if there is a change, it is small.

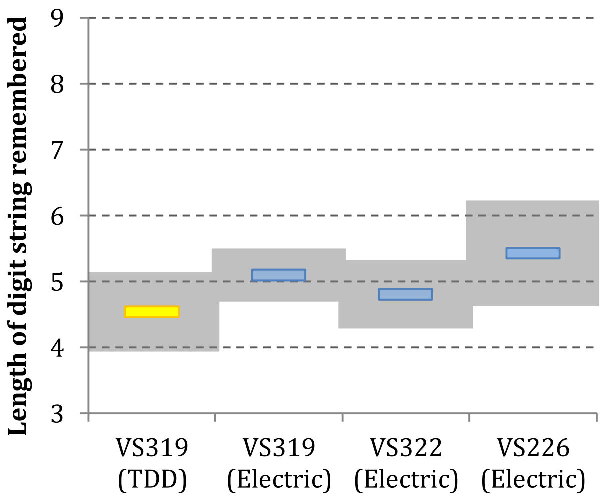


Figure 4-17: Mean DSB scores in the different situations. Coloured markers show the average, shaded areas show the margin of error (95% confidence).

		TDD	ELECTRIC		
		VS319	VS322	VS226	VS319
		n 24 m 4.54 sd 1.22	*n* 30 m 4.81 sd 1.20	*n* 25 m 5.43 sd 1.67	*n* 45 m 5.10 sd 1.18
TDD	VS319		dif. 0.27 t 0.80 p ns	dif. 0.89 t 2.13 p <0.05	dif. 0.56 t 1.83 p <0.1
ELECTRIC	VS322			dif. 0.62 t 1.56 p ns	dif. 0.29 t 1.04 p ns
	VS226				dif. -0.33 t -0.88 p ns
	VS319				

Figure 4-18: Significance of the differences between the DSB scores in the different situations and associated statistics. Note: significant differences are highlighted green, marginally significant ones are blue, while non-significant ones are grey.

However, the paired comparisons challenge this (Table 4-13). The change in individuals' scores suggests that, on average, there was no real change in performance. It suggests that the apparent change might have been due to a number of less high-performing subjects participating in the second round, but not the first, or that the effect is not reliable.

	n	Mean difference	SD	t	significance
VS319: TDDs vs. Electric	18	-0.06	1.03	-0.23	ns
VS322 & 226	18	-0.06	1.53	-0.15	ns

Table 4-13: Working memory: Round 1 vs. Round 2: Paired comparison using the differences in individuals' scores in each situation. Control groups have been combined to make up for their low sample size.

Breaking down the results by test round raises further doubts about the effect (Figure 4-19). As shown below, VS322 shows a similar decrease in performance in the second round — suggesting that scores were reduced by some external factor in the second round. A possible explanation could be because they were more distracted as the test was on the first tutorial back after the mid-semester break. While VS226 does not also show a reduction, it should be noted that its variability in test scores was much higher, making its results less reliable. Another possibility is that people were less interested in doing the test again, and so didn't try as hard.

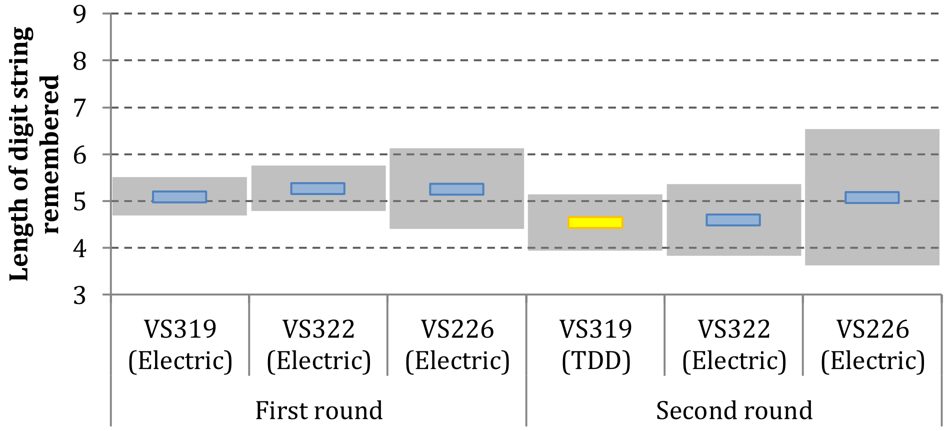


Figure 4-19: Average scores of working memory tests broken down by room and testing round. Shaded areas show the margin of error (95% confidence)

Inside either round, none of the rooms were significantly different (Figure 4-20,Figure 4-21). Both VS319 (dif. = -0.56, t(67) = 1.83, p<0.1) and VS322 (dif. = -0.68, t(39) = 1.82, p<0.1) had *marginally* significant lower average scores in the second round. This suggests that the TDDs were not having any significant impact on performance, as the same effect was found in the control room.

	VS319		VS322		VS226	
	n	45	*n*	20	*n*	21
	m	5.10	m	5.28	m	5.26
	sd	1.18	sd	0.90	sd	1.62
VS319			dif.	-0.18	dif.	-0.16
			t	-0.66	t	-0.41
			p	ns	p	ns
VS322					dif.	0.01
					t	0.03
					p	ns

Figure 4-20: First round of Digit Span testing: Significance of the differences between scores in the different rooms and associated statistics.

	VS319		VS322		VS226	
	n	24	*n*	21	*n*	15
	m	4.54	m	4.60	m	5.08
	sd	1.22	sd	1.45	sd	1.94
VS319			dif.	-0.05	dif.	-0.54
			t	-0.13	t	-0.97
			p	ns	p	ns
VS322					dif.	-0.49
					t	-0.82
					p	ns

Figure 4-21: Second round of Digit Span testing: Significance of the differences between scores in the different rooms and associated statistics.

To look at this more closely, a linear regression model was made from the DSB scores to analyse the effect these factors — the different rooms, rounds, and the TDDs — had on the average performance. The statistical significance of the effect of each of the factors is presented below (Table 4-14):

	t	p-value
VS226	0.81	0.42
VS322	0.20	0.84
2nd round	-1.47	0.14
TDDs	-0.20	0.84

Table 4-14: Working memory: Significance of various factors produced by the regression model examining the difference of TDDs to the mean performance controlling for effects of room and round. Full output in Appendix F2.

The model shows that, when all the factors are controlled for, none of them have a *significant* impact on the average performance score. The factor that has the biggest impact, and is closest to being significant, is the effect of *round*, with a p-value of 0.14. This corroborates the previous analysis, and shows that the possible negative effect of TDDs that was found is most likely a small round effect from some external factor, and not related to the TDDs.

4.1.2.2 Processing speed

The processing speed scores here are the average number of letters the subjects read in four seconds.

The results show no significant differences between the rooms/conditions (Figure 4-22,Figure 4-23). This is corroborated by the paired comparison, which also found no significant change in score between tests (Table 4-15).

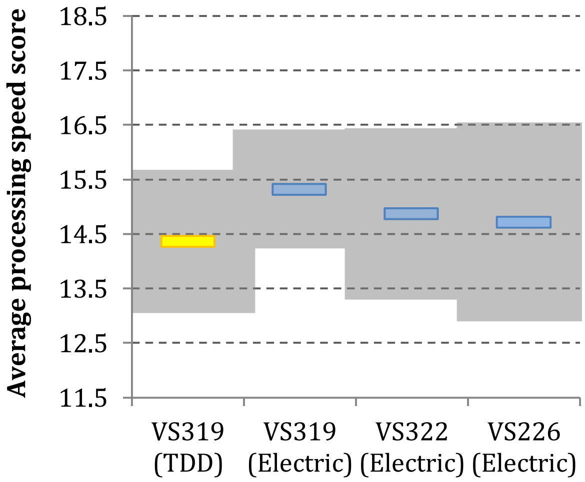


Figure 4-22: Mean processing speed scores in the different situations.

		TDD	ELECTRIC		
		VS319	VS322	VS226	VS319
		n 23 m 14.36 sd 2.62	*n* 24 m 14.87 sd 3.20	*n* 24 m 14.72 sd 3.72	*n* 41 m 15.32 sd 3.00
TDD	VS319		dif. 0.51 t 0.64 p ns	dif. 0.36 t 0.39 p ns	dif. 0.96 t 1.38 p ns
ELECTRIC	VS322			dif. -0.15 t -0.16 p ns	dif. 0.46 t 0.62 p ns
	VS226				dif. 0.61 t 0.70 p ns
	VS319				

Figure 4-23: Significance of the differences between the processing speed scores in the different situations and associated statistics.

	n	Mean difference	SD	t	significance
VS319: TDDs vs. Electric	21	-0.50	2.03	-1.13	ns
VS322 & 226	20	0.07	2.77	0.12	ns

Table 4-15: Processing Speed: Round 1 vs. Round 2: Paired comparison using the differences in individuals' scores in each situation. Control groups have been combined to make up for their low sample size.

Breaking down the results by round shows nothing new (Figure 4-24). There are no significant differences between any of the rooms or test sessions.

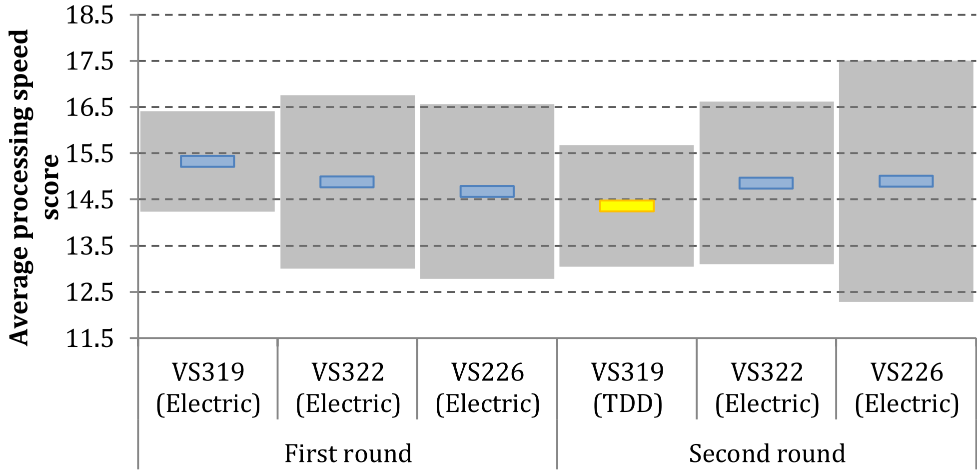


Figure 4-24: Average scores of processing speed tests by room and testing round.

Interestingly, as with the DSB scores, the scores under TDDs are the lowest of all the groups. However, there is no significant evidence that the lighting is having any impact on processing speed.

It should also be noted that the Architecture students in VS319 had a hand-in in another course the day after the first round. However, with the absence of any significant differences between any of the rooms there is no evidence that this affected their performance.

4.1.2.3 Discussion

Overall, it is concluded that there is no evidence that TDDs have any effect on cognitive performance.

The only indications of any effects were negative and it was most likely one of *round* rather than TDDs. The fact that TDDs had the lowest scores in both tests *may* indicate a possible trend — however the lack of significance means that it cannot be substantiated, and it may just be a coincidence.

The lack of performance effects, while disappointing, is not too surprising. As discussed in the Methodology, most studies examining performance effects of lighting have had only limited success. This does provide an interesting contrast with the Heschong Mahone Group study (2003), which found a correlation between amount of daylight and performance on working memory tests. The TDDs provide similar levels of daylight, so the lack of effect here may be due to limited exposure. The HMG study involved subjects who were working all day in the conditions being investigated. Other studies have suggested that daylighting may affect health and stress (e.g. Çakir & Çakir, 1998), which could impact performance, but would require more long-term exposure.

The lack of performance effects may also be because the tests did not have enough power to detect them. While the paired t-test was powerful enough to have good odds of detecting an effect of the same size as the Heschong Mahone study (80% chance of detecting a difference of 0.7 on the digit span test), it has a much lower probability of detecting more subtle differences. It is unfortunate that participation was not maintained better in the second half of the semester, as it would have provided more power.

The findings do not rule out the possibility that there could be performance benefits — especially in people working under TDDs all day. However, they do suggest that it is unlikely that effects will appear in short timeframes, and that a significantly larger or longer study is needed to uncover such effects.

4.1.3 Mood

Mood was measured three times in the study. To deal with its high variability, it was planned to assess it multiple times — at least twice in each half of the semester. Unfortunately, participation completely dropped off later in the second half of the semester, with only four subjects completing the fourth mood test, so there was only one survey completed under TDDs. Thus, there are only three mood tests in total.

As discussed in the methodology (3.7.3), two facets of mood are measured: pleasure, and arousal.

Mood was assessed as *change* in mood, as single measures of mood are likely to be heavily influenced by how the subjects felt *before* they came to tutorial, and precedent suggests that such measures are unlikely to find any effects (Boyce et al., 2003).

4.1.3.1 Change in pleasure

The results show that under TDDs, the students' had a significantly more positive change in pleasure during their tutorial (Figure 4-25,Figure 4-26). This suggests that the TDDs may have encouraged the students in VS319 to feel better during the tutorial. The paired comparison also supports this apparent effect (Table 4-16).

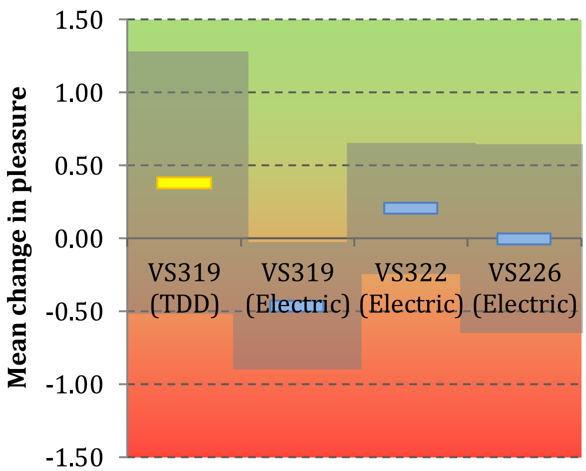


Figure 4-25: Mean change in pleasure in the different situations.

		TDD	ELECTRIC		
		VS319	VS322	VS226	VS319
		n 14, m 0.38, sd 1.33	n 22, m 0.21, sd 0.88	n 26, m 0.00, sd 1.38	n 22, m -0.46, sd 1.23
TDD	VS319		dif. 0.23, t 0.55, p ns	dif. 0.44, t 1.03, p ns	dif. 0.90, t 2.56, p <0.02
ELECTRIC	VS322			dif. 0.34, t 0.74, p ns	dif. 0.80, t 3.07, p <0.005
	VS226				dif. 0.46, t 1.55, p ns
	VS319				

Figure 4-26: Significance of the differences between the changes in pleasure in the different situations and associated statistics.

	n	Mean difference	SD	t	significance
VS319: TDDs vs. Electric	14	0.60	0.84	2.67	p<0.02

Table 4-16: TDDs vs. electric lighting in VS319: Paired comparison of changes in pleasure using the differences in individuals' results in each situation.

Breaking the results down by round, however, argues against the apparent effect (Figure 4-27). Note that very few of the students in VS322 completed the surveys in the second test (n = 4), as they had an assignment in another course due that day.

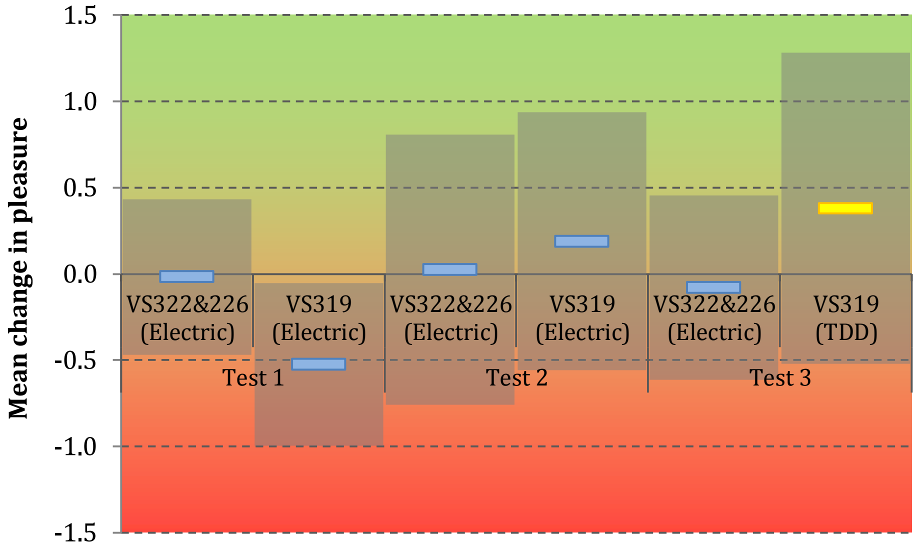


Figure 4-27: Mean changes in pleasure over the course of the tutorials broken down by test round. Control groups combined to make it easier to see the key points.

Looking at the results (Figure 4-27), it can be seen that the change in pleasure under TDDs is *not* significantly different to what it was in the *second* round under electric lighting (dif. = 0.19, t(35) = 0.4, ns), nor is it significantly different to the control groups (Figure 4-26). While change in pleasure under TDDs was significantly more positive under TDDs than it was under electric lighting *in the first round*, the same is true when comparing the first round in VS319 to the second round, or to the control rooms (Table 4-17). The apparent effect of TDDs is most likely due to the fact that the students in VS319 were having a "bad day" during the first test round, with a significantly more negative change in mood during the tutorial than normal. Given that in the second round, which had the same electric lighting, the results were significantly more positive, the "bad day" would appear to be unrelated to the lighting.

VS319 round 1 vs:	*n*	Mean difference	t	significance
VS319 (TDDs)	38	0.91	2.22	p<0.05
VS319 round 2 (Electric)	57	0.71	1.92	p<0.1
Control rooms round 1	62	0.51	1.83	p<0.1
Control rooms round 2	56	0.55	1.81	p<0.1
Control rooms round 3	55	0.45	1.80	p<0.1

Table 4-17: Statistical comparisons of how the results for change in pleasure in VS319 in round 1 differ from the result for all the other rooms and rounds. While most are only at marginal significance, the consistent overall trend is clear.

4.1.3.2 Change in arousal

Comparisons of the average changes in mood (Figure 4-28) show no significant differences between the change in arousal under TDDs and under electric lighting (Figure 4-29). The only point of note is that the students in VS322 were apparently more stimulated during their tutorials than the others, with a more positive change in arousal.

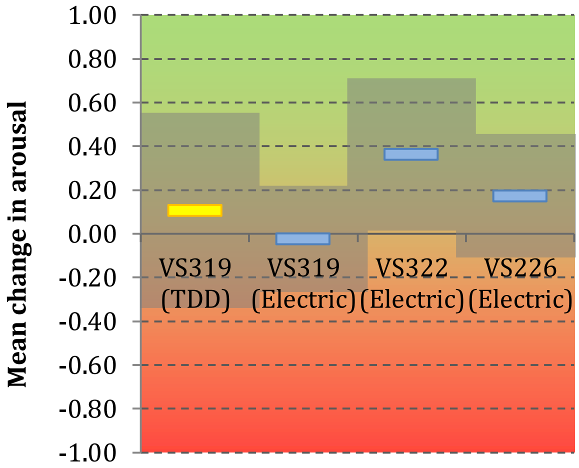


Figure 4-28: Mean change in arousal in the different situations.

		TDD	ELECTRIC		
		VS319	VS322	VS226	VS319
		n 14 m 0.38 sd 1.33	*n* 22 m 0.21 sd 0.88	*n* 26 m 0.00 sd 1.38	*n* 22 m -0.46 sd 1.23
TDD	VS319		dif. 0.23 t 0.55 p ns	dif. 0.44 t 1.03 p ns	dif. 0.90 t 2.56 p <0.02
ELECTRIC	VS322			dif. 0.34 t 0.74 p ns	dif. 0.80 t 3.07 p <0.005
	VS226				dif. 0.46 t 1.55 p ns
	VS319				

Figure 4-29: Significance of the differences between the changes in arousal in the different situations and associated statistics.

However, the paired comparison suggests a possible effect of TDDs, with change in arousal being more positive at marginal significance (Table 4-18).

	n	Mean difference	SD	t	significance
VS319: TDDs vs. Electric	14	0.56	1.00	2.11	p<0.1

Table 4-18: TDDs vs. electric lighting in VS319: Paired comparison of changes in arousal using the differences in individuals' results in each situation.

Breaking the results down by round shows nothing new, and the results under TDDs are no different from those under electric lighting (Figure 4-30).

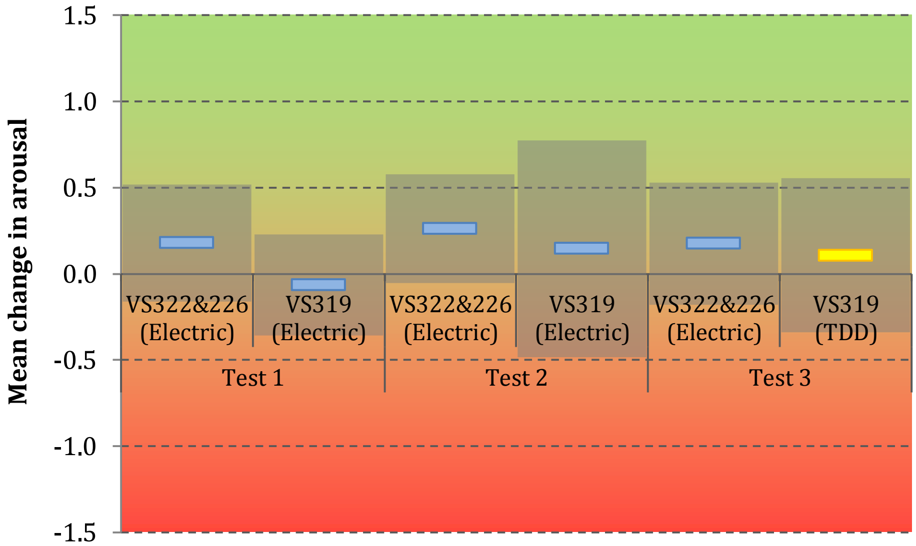


Figure 4-30: Mean changes in arousal over the course of the tutorials in the different rooms and sessions.

So, only the paired comparison suggests any effect — and it is only at *marginal* significance. The paired comparison can usually be considered to be more sensitive, as by looking at the changes in an individual's responses rather than the difference between the means of two groups, we can control for the effects of the differences between people. In this case, however, it may not be a very good measure, as correlations between the changes in arousal on the different days are poor, lacking both significance and consistency (Figure 4-31). This suggests that a person's mood on one day is not a good predictor of their mood on another — possibly due to the variation in their tutorial experiences — and thus that there is no real benefit from trying to control for the individual differences. Indeed, given that the paired comparison has a lower total sample size, its results could be misleading.

	Round 2	Round 3
Round 1	0.32 (n = 26)	0.13 (n = 21)
Round 2		-0.28 (n = 17)

Figure 4-31: Correlations of measures of change in arousal between the different rounds. None are significant (p>0.1).

Overall, the evidence does not support an effect of TDDs on arousal.

4.1.3.3 Discussion

Unfortunately, the tests have several issues that prevent the results from being able to say anything conclusive. Specifically:

1) The small sample size of the later tests.
2) The variability of the tutorial experience confounding the results.
3) Not enough tests being completed to provide sufficient data-points.

4.1.3.3.1 Small sample size

The sample size of the groups, especially TDDs, were really too small relative to what was called for in the original research design. Participation rates in the mood surveys were always lower than those in the other tests, because many people left the tutorial without finishing the second survey. When participation dropped in the second half of the semester, it left only 14 subjects in the group under TDDs. This reduces the reliability of the results, and makes it harder to extrapolate them to the broader population. The low statistical power means that only reasonably large effects would be likely to be found. However, mood effects that have been found in lighting studies are generally small — for example, an difference of ~0.3 (Newsham et al., 2004) — and the sample sizes here provide only about a 10-20% chance of detecting an effect of that magnitude.

4.1.3.3.2 Variation in task and stimulus

A bigger problem is effects of the variation in the groups' experiences in the different tutorials. The "bad day" that the group in VS319 in the first test had is a prime example of this — the tutorials that group had that day had a largely negative impact on their mood. This uncontrollable variation increases the variability of the results, making it harder to identify subtle differences.

To deal with the effect of the variation here, the tests need to be run multiple times so that the variation in tutorial experience can be averaged out. This leads into the biggest problem:

4.1.3.3.3 Not enough data-points

Because of the variation in mood effects caused by differences in tutorial experiences, the tests need to be run multiple times. This is why, unlike the other tests, the mood surveys were run twice in the first half of the semester. It was intended that the same would be done in the second half. Unfortunately, virtually no-one completed the surveys in the fourth round (n=4). Hence there is only one set of mood surveys under TDDs. The problem is that with only one test it is impossible to say with confidence whether or not it was a relatively "good" or "bad" day. It is possible, for instance, that the tutorial experience that day under TDDs would have been a negative one like the "bad day" — but that because of the TDDs it became positive. Without multiple tests there is no way of addressing the "tutorial" component of the mood, and no real conclusions can be drawn about whether or not the lighting had any effect.

4.1.3.3.4 Summary: tests unsuccessful

So, to sum up the results of the mood tests:

1) There is no good evidence of any effects of TDDs on mood.
2) There are not enough data-points to deal with the confounding effects of the tutorial experience, making it difficult to draw conclusions either way.
3) The sample size of the TDDs group is too small.

Therefore, while no effects of TDDs on mood were found, it cannot be concluded that there are no effects of TDDs on mood.

4.1.4 Sleepiness

Subjective sleepiness was measured at the beginning of every test, and describes the average sleepiness over a more substantial number of tutorials. While it may be argued that the short exposure time gives little chance for the lighting to affect people, the study by Lehrl et al. (2007) has suggested that light spectrum *could* have small short-term alerting effects.

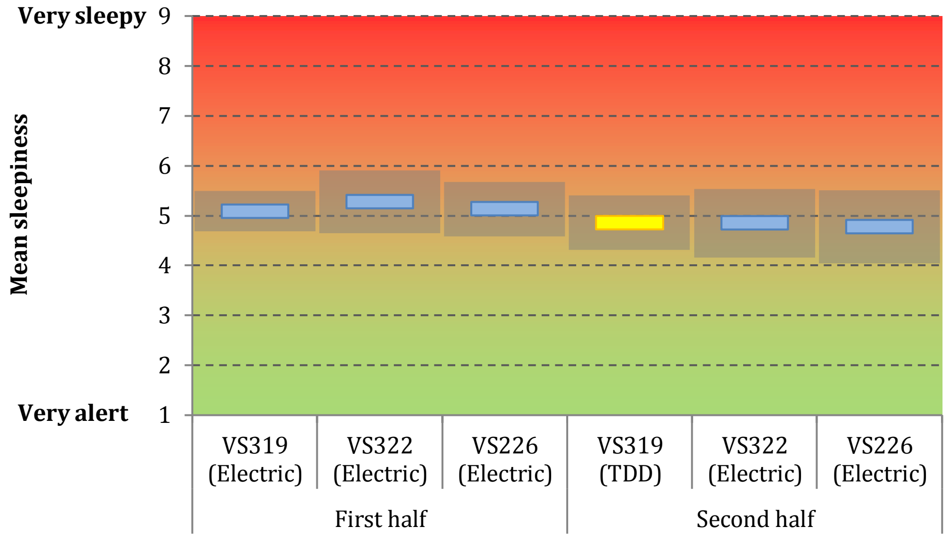


Figure 4-32: Comparison of average sleepiness in the first and second halves of the semester by room

The results above show, fairly clearly, that the TDDs did not cause any significant difference in mean sleepiness in students in that room (dif. = -0.23, t(104) = 0.79, ns). Indeed, the average sleepiness is very consistent. Figure 4-33 shows that there is some variation, but it is mainly in VS226, which likely due to its changing tutorial times — the "peaks" there occur on days when the sample was primarily from the afternoon class, showing effects of time and circadian rhythm.

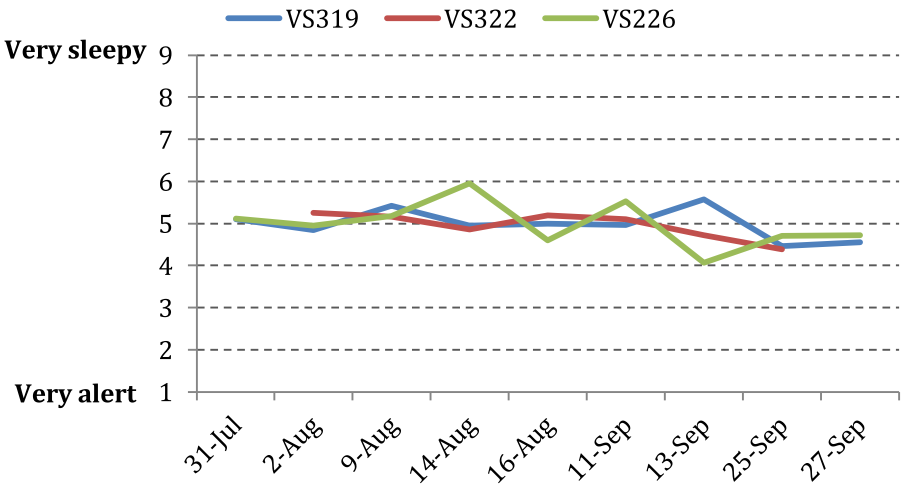


Figure 4-33: Mean sleepiness in each room across all tests

The lack of any overall effect may be due to the confounding influence of individual factors such as variation in sleep quality. Lehrl et al. (2007) assessed the different lighting conditions one after another in the same session on the same people, and so these factors were less of an issue.

4.2 Could other (non-lighting) factors be influencing the results?

To be confident that the observed effects are due to the lighting, it is necessary to rule out alternative explanations. It is also possible that other factors could be covering up effects of the lighting and preventing them from being found. Thus, a number of other environmental and demographic factors were analysed to assess their influence, and control for their effects where appropriate. Factors analysed were:

1) Air temperature
2) CO_2 levels
3) Noise levels
4) Subjects' gender
5) Whether or not English was the subjects' native language
6) Whether or not the subjects' wore glasses
7) Whether or not they had eaten beforehand
8) Whether or not they had drunk coffee beforehand
9) Sleepiness

4.2.1 Environmental conditions

4.2.1.1 Air temperature

There are three key points to be seen in the graphs below:

1) The temperatures in the different rooms and tests are generally within about 2-3°C of each other.
2) They are mostly inside the typical "comfortable" temperature range of 20-24°C (NZS 4243).
3) The lone outlier is VS226 in the afternoon, which has a tendency to overheat, rising above 25°C.

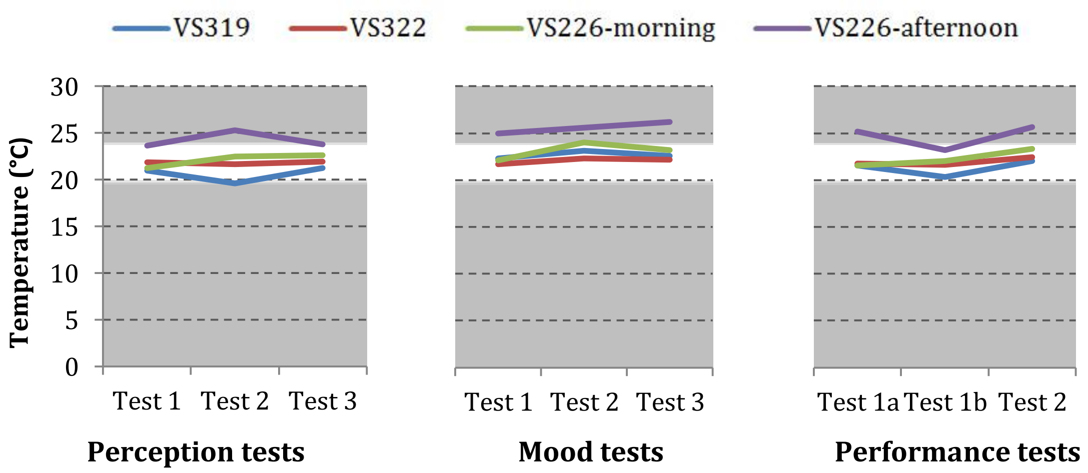


Figure 4-34: Air temperatures in rooms during the different tests. Note: as the students had difficulties following the instructions the first time, the processing speed test had to be rerun in the first half of the semester. Test 1a is the first Digit Span test, Test 1b is the first processing speed test. White band shows comfort range (NZS 4243).

Given this, it is unlikely that the temperature is having any significant impact upon most of the results. VS226, while being the warmest room and most different from the others, has not given significantly different results to the other artificially lit rooms.

A lack of noticeable effects is consistent with other research. Studies looking at the effects of temperature on performance generally look at larger temperature differences than this (e.g. Hygge & Knez, 2001), and effects on simple mental tasks have been variable and not necessarily apparent (Oseland, 1999). Meta-analysis by Seppänen et al. (2006, in Leyten et al., 2012) suggests that performance is optimum between 20-24°C, and that the ~25°C temperatures in VS226 would only reduce performance by ~2-3% - which would not be noticeable in this study, especially as only a small portion of the total sample is outside the "good" temperature range. Moreover, effects tend to be lessened when exposure times are short (Leyten et al., 2012), which would further reduce the likelihood of any effects.

Additionally, any effects would be limited to the comparisons with the control group in VS226, and would not affect the comparisons between TDDs and electric lighting in VS319, or the comparisons with the other control group.

So, while effects of temperature on subject response cannot be ruled out, it is unlikely that temperature differences are significantly affecting the results.

4.2.1.2 Air quality

The graph below shows several things:

1) CO_2 levels in the rooms vary. Variance in a room is generally within about 300ppm.
2) There are some clear trends: VS319 has the lowest CO_2 levels, VS226 is generally about 200-300ppm higher, and VS322 has the highest levels.
3) Both VS319 and VS226 stay under the limit of 1000ppm recommended by NZS 4303 for acceptable indoor air quality.
4) VS322 is over the recommended limit, especially by the second tutorial, where it can reach 1400ppm.

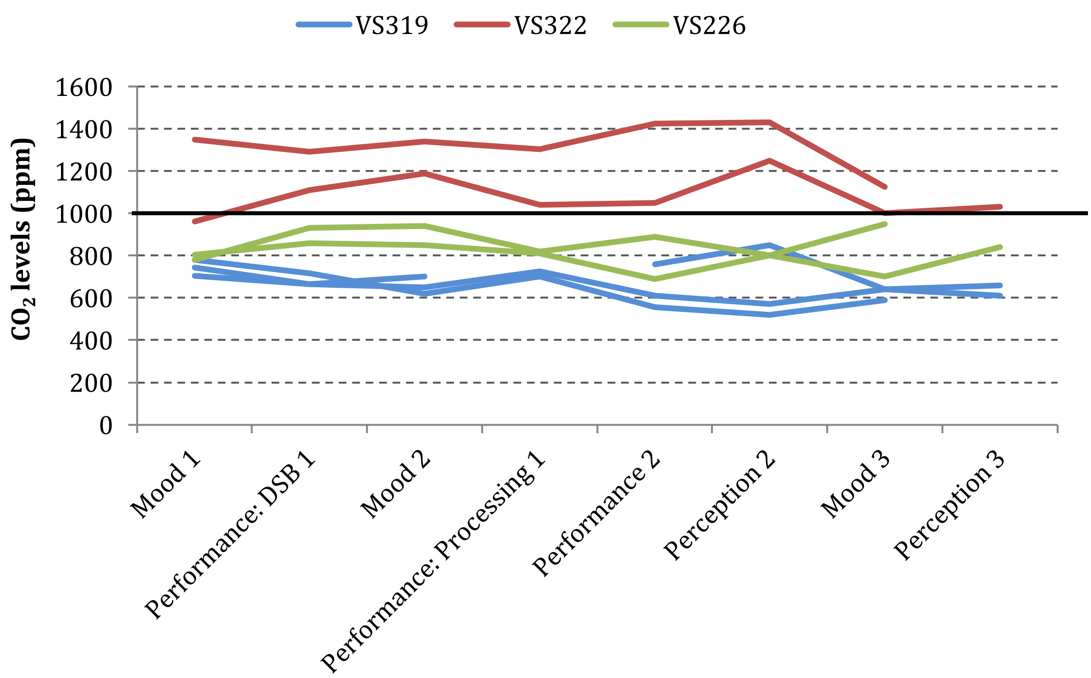


Figure 4-35: CO_2 levels in the different rooms across the test sessions and tutorials. Note: The multiple lines for each room are following the different tutorial streams in the rooms. There are two tutorials each in VS322 and VS226, and three in VS319. The black line at 1000ppm marks the limit recommended by (NZS 4303).

Studies have suggested CO_2 levels around those in VS322 can negatively affect performance (e.g. Wargocki & Wyon, 2012). This raises the possibility that performance in VS322 is being reduced by its elevated CO_2 levels, and that without this factor, performance would actually be higher relative to the other rooms.

There is no compelling reason, however, to believe that this is the case here. VS322 has not evinced any systemic trend of lower (or higher) performance, compared to the other rooms .

It cannot alter the conclusions about mood due to the test's problems — although it is interesting to note that VS322 did consistently trend higher than the other rooms on *change in arousal* — which could almost suggest that the CO_2 was helping to stimulate the subjects. Even if that was the case, however, controlling for it would only result in VS322's results becoming closer to that of the other rooms.

Overall, differences in air quality do not appear to be significantly affecting the results.

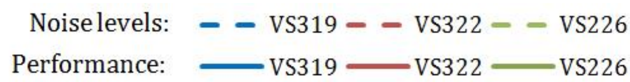
4.2.1.3 Noise

Noise measurements were taken during the performance tests, as noise is potentially a distractant and annoyance that could reduce performance (Oseland, 1999).

The graphs below show how noise levels measured in the rooms during the tests relate to the average test scores (Figure 4-36,Figure 4-37).

They show:

1) That noise levels were higher during the second round - possibly due to the students having just returned from their mid-semester break.
2) A possible trend with the working memory scores, with scores consistently lower in the second round, while noise levels were consistently higher.
3) No consistent effect of noise on processing speed results — while VS319 had lower scores in the second round, the other rooms did not.



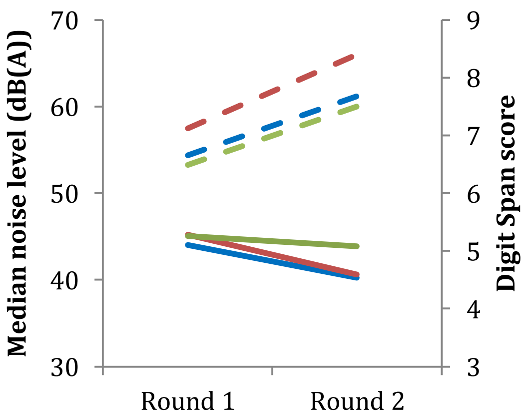


Figure 4-36: Median noise levels compared to mean scores on working memory tests

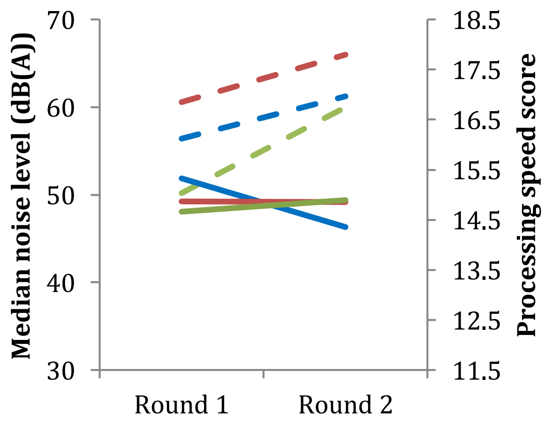


Figure 4-37: Median noise levels compared to mean scores on processing speed tests

A lack of any clear effects of noise on processing speed is not unsurprising, as the noise levels found here are not atypical for offices and classrooms (European Agency for Safety and Health at Work, 2005), and past research into its effects has had inconsistent results, with positive, negative and null results found (Oseland, 1999).

The comparison with the working memory scores suggests that the noise levels may be connected to the possible *round* effect that was found (section 4.1.2.1), with the increased noise levels reducing performance.

Explaining the round effect would not, however, change the overall results. There were no significant differences between the different rooms in either round. Any attempt to control for noise effects would not be expected to change that, as the increase in noise between rounds was roughly the same across all three rooms, and so it would not change the fact that the TDDs were no different from the rooms with electric lighting.

4.2.2 Subject characteristics

The subject characteristics discussed here are broken down into two groups: 1) subject demographics and 2) pre-test conditions of the subjects.

To see if they could be influencing the results, three questions were asked:

1) Is the sample makeup significantly different between rooms/sessions?
2) Is there evidence that the characteristic has any significant effect on the results (e.g. do females give different responses to males?)
3) If there could be an effect, does controlling for it change the conclusions drawn from the results?

4.2.2.1 Could differences in subject demographics between the rooms change the results?

This question looks specifically at comparisons between the test room and the control rooms. Any variation in demographic proportions between test sessions is already controlled for by the paired comparisons.

This section examines gender, English being a second language, and the wearing of glasses. Age and colour-blindness were not examined, because almost all of the subjects were of the same age group, and none reported being colour-blind.

A series of comparisons and statistical tests were run to examine the possible effects of the various factors. They are shown in full and explained in the assessment of the influence of gender, below. Later sections only report the key findings and statistics, for brevity's sake. The full statistical summaries may be found in Appendix G.

4.2.2.1.1 Gender

To find out whether or not gender (or other factors) could be affecting the results, the first question to ask is whether or not the proportions actually differed between the rooms. Whether or not gender plays a role in lighting effects does not matter if there are no gender differences between groups. The demographics of the groups in each of the tests are shown in the graphs on the next page (Figure 4-38).

In the graphs, we see that the demographics vary across the rooms, with VS319 generally having a lower proportion of female subjects, and the largest difference being about 40 percentage points. This means that there is the *potential* for variation in gender ratios to bias the results.

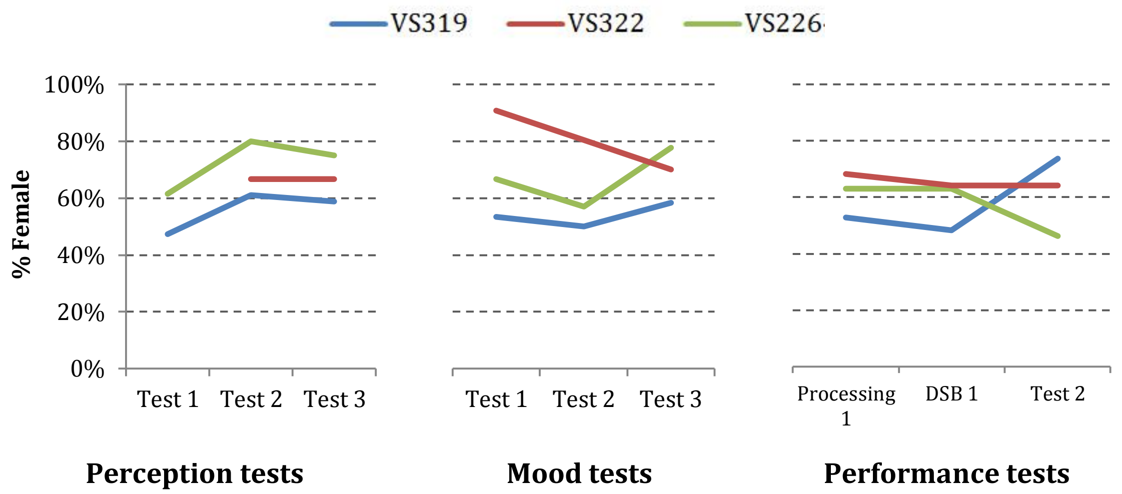


Figure 4-38: Proportion of sample that is female in the different rooms and tests

Accordingly, the test results were broken down by gender. The analysis only uses results from under electric lighting in order to avoid possible confounding effects from the TDDs. The possibility that the different genders could respond to the TDDs differently was not examined in this study as it was outside its scope, and the samples under TDDs were too small to split.

	Male: *n* = 27		Female: *n* = 42				
	Mean	SD	Mean	SD	Mean difference	t	significance
Lighting Quality	2.35	0.78	2.30	0.65	-0.05	-0.30	ns
Glare Problems	1.39	1.09	1.41	1.15	0.02	0.08	ns
Daylight Quality	0.12	0.43	0.25	0.62	0.13	1.04	ns
Room Attractiveness	3.63	1.22	3.52	1.36	-0.11	-0.35	ns
Illumination	4.29	1.16	4.33	1.31	0.04	0.14	ns

Table 4-19: Gender differences in perception of lighting using the results of the groups under electric lighting.

	Male			Female					
	n	Mean	SD	*n*	Mean	SD	Mean difference	t	significance
DSB	27	3.26	0.97	36	2.99	1.30	0.27	0.96	ns
Processing	24	15.72	3.62	34	13.93	2.61	1.79	2.08	p<0.05

Table 4-20: Gender differences in performance tests using the results of the first round of testing under electric lighting.

	Male			Female					
	n	Mean	SD	*n*	Mean	SD	Mean difference	t	significance
Δ Pleasure	31	0.08	1.27	48	-0.37	1.10	0.45	1.62	ns
Δ Arousal	31	0.10	0.58	48	-0.04	0.75	0.14	0.93	ns

Table 4-21: Gender differences in change in mood using the results from the first two surveys under electric lighting.

	Male			Female			Mean difference	t	significance
	n	Mean	SD	*n*	Mean	SD			
Sleepiness	44	4.98	1.35	54	5.17	1.45	0.19	0.66	ns

Table 4-22: Gender differences in sleepiness using the results from the first half of the semester under electric lighting.

Comparison of gender averages reveals that the differences between the male and female responses for perception of the room (Table 4-19), change in mood (Table 4-21), working memory scores (Table 4-20), and sleepiness (Table 4-22) are not significant. Females do, however, have significantly lower average **processing speed** scores.

To determine whether or not variations in gender ratio are actually changing the results, the scores were adjusted to those of an equivalent all-male sample using the difference between male and female averages. This controls for the effects of gender, and provides an estimate of what the differences between rooms would be like if gender didn't vary (Figure 4-39).

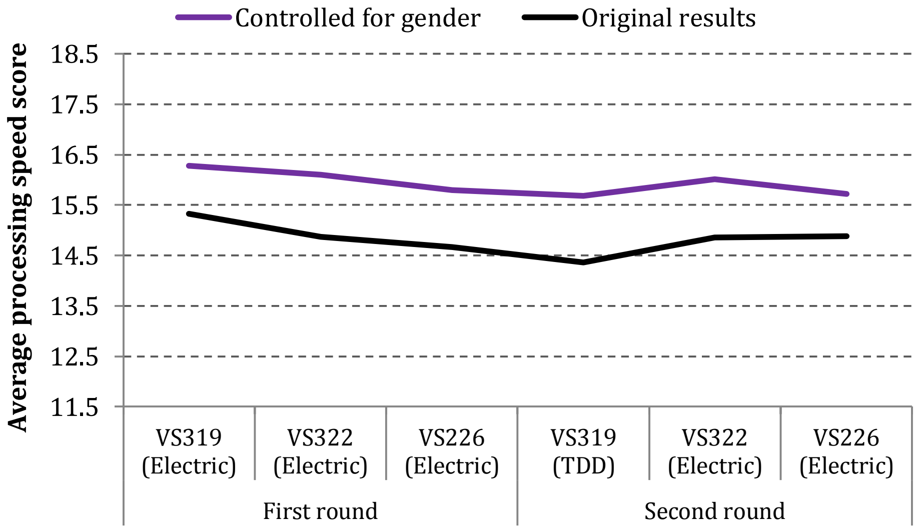


Figure 4-39: Processing speed scores adjusted to control for gender effects compared to original results

As can be seen in the graph (Figure 4-39), controlling for gender does not change the comparisons between the groups — there are still no significant differences between them. Indeed, if anything, it actually reduces the variation. Thus, we can conclude that variations in gender ratios did not have any significant impact on the results.

4.2.2.1.2 English as a Second Language

Comparing the responses of the native English speakers to the non-native shows that there were no significant differences between them — and also that for the most part the number of ESOL students is relatively small (*n*: 10-17). The only difference approaching significance is in **perceived illumination** (dif. = 0.63, t(67) = 1.47, ns). Controlling for it, by converting the results to those of an equivalent all-native-English-speaking group, has no significant effect on the results, as shown below:

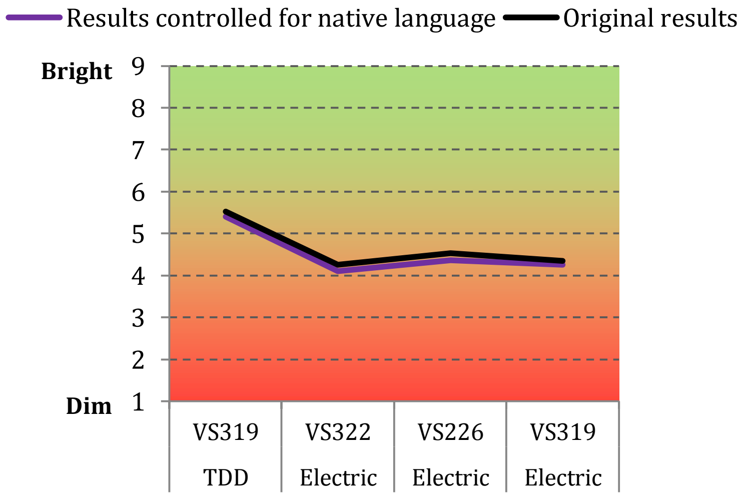


Figure 4-40: Illumination scores adjusted to control for possible effects of subjects' not being native English speakers compared to original results

4.2.2.1.3 Wearing Glasses

Comparing the results of those who do or do not wear glasses shows that there are generally no significant differences between them. Only the **working memory** test has a difference close to being meaningful (dif. = 0.47, t(61) = 1.54, ns), and as shown below, controlling for it makes no difference to the results.

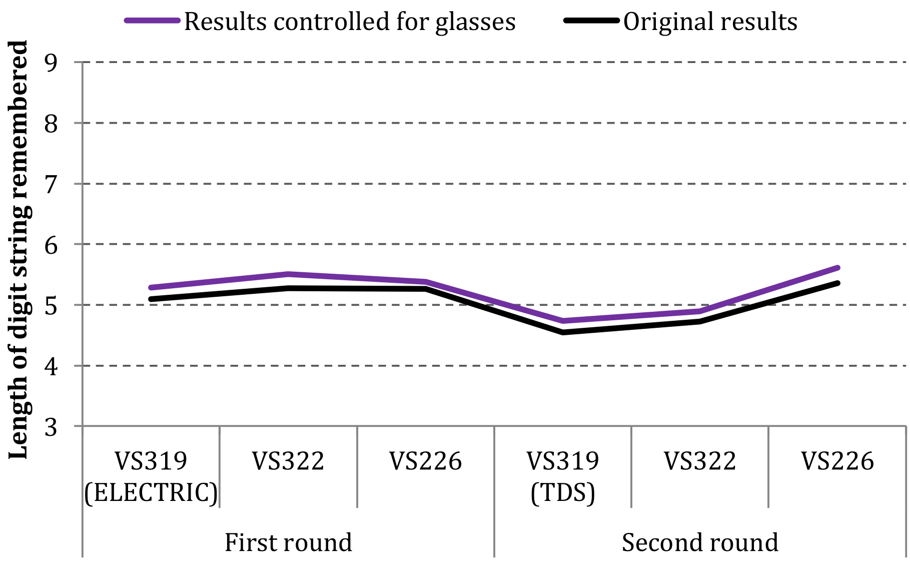


Figure 4-41: Digit Span scores adjusted to control for possible effects of wearing glasses compared to original results

4.2.2.2 Could differences in the pre-test conditions of the subjects be affecting the results?

Unlike the demographics, the pre-test conditions of the subjects — such as if they've drunk coffee — can change between test sessions. Thus, if the pre-test conditions of the subjects are influencing the results then they could affect both the comparisons between rooms, and the comparisons between test sessions. This means that it is also necessary to control for effects on the paired comparisons of the individuals' change in response between sessions. This is done by adjusting the subjects' results by the difference linked to the factor (e.g. drinking coffee) to convert all the subjects to the same condition (e.g. have all drunk coffee), before running the paired comparisons.

Again, the full statistics are in Appendix G.

4.2.2.2.1 Whether or not they'd eaten breakfast/lunch

Comparisons of the results of those who had eaten or had not eaten found significant differences between the groups with regards to **lighting quality** (dif. = 0.42, $t(85) = 2.36$, $p<0.05$), **room attractiveness** (dif. = 0.85, $t(85) = 2.71$, $p<0.01$), and **sleepiness** (dif. = 0.93, $t(85) = 2.23$, $p<0.05$). **Glare problems** (dif. = 0.44, $t(85) = 1.47$, ns) and **illumination** (dif. = 0.44, $t(85) = 1.41$, ns) have differences large enough that they might be able to have an effect, even if they are not significant. Students that had eaten beforehand were less sleepy and tended to give lower ratings, meaning they were more critical about lighting quality and room attractiveness — but interestingly enough *less* negative about glare problems, where lower scores are better (although the difference there did not achieve significance).

There were no significant differences in performance scores, or change in pleasure. There was a marginally significant difference in change in **arousal** (dif. = 0.29, $t(96) = 1.81$, $p<0.1$), with students that haven't eaten tending to have a more negative change in arousal over the course of the tutorial. This makes intuitive sense, as one would expect hungry students to be more prone to getting tired and less alert, and it matches the effect found on sleepiness.

As shown below, controlling for having eaten beforehand does not change the results significantly (Figure 4-42). The paired comparisons between the TDDs and electric lighting in VS319 also remained unaffected (Table 4-23), as did the comparisons between the sunny and overcast days (Table 4-24).

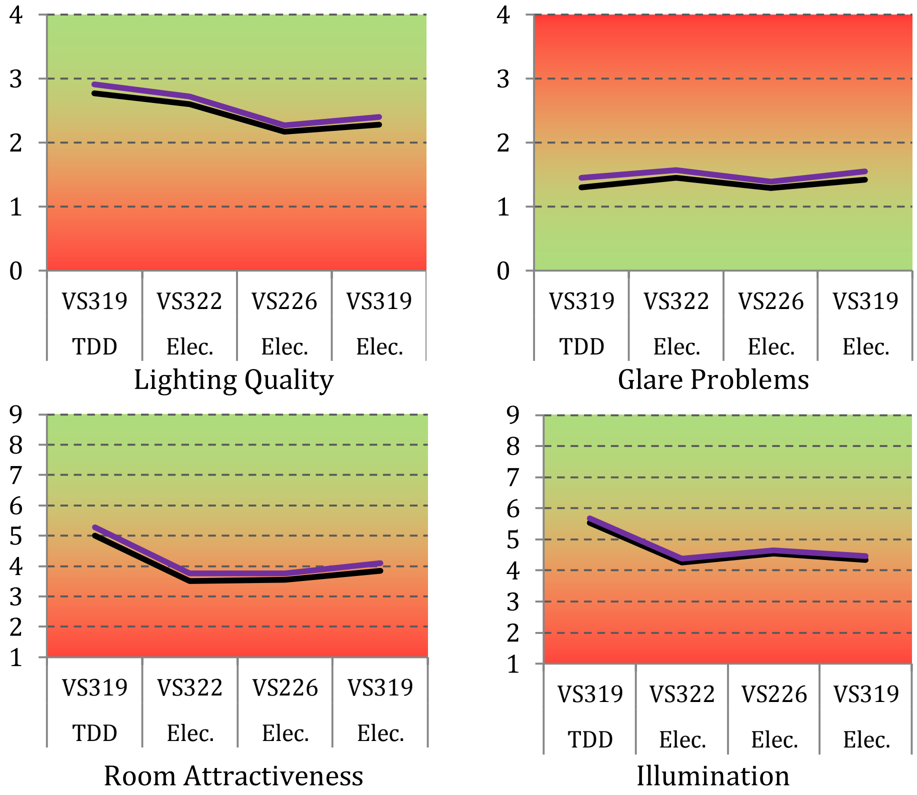


Figure 4-42: Perception results controlled for having eaten.

	Mean difference	SD	t(24)	significance	Original significance
Lighting Quality	0.78	0.97	4.01	p<0.005	p<0.005
Glare Problems	-0.05	0.93	-0.28	ns	ns
Room Attractiveness	1.50	1.95	3.84	p<0.005	p<0.005
Illumination	1.18	1.49	3.96	p<0.005	p<0.005

Table 4-23: TDDs vs. Electric lighting in VS319: Paired comparison using the differences in individuals' responses in each situation with responses controlled for having eaten.

	Mean difference	SD	t(15)	significance	Original significance
Lighting Quality	0.01	0.84	0.05	ns	ns
Glare Problems	0.51	0.93	2.20	p<0.05	p<0.05
Room Attractiveness	-0.05	1.25	-0.15	ns	ns
Illumination	0.56	1.05	2.15	p<0.05	p<0.05

Table 4-24: Overcast vs. sunny day in VS319: Paired comparison using the differences in individuals' responses in each situation with responses controlled for having eaten.

Controlling for having eaten also does not change the conclusions about sleepiness. As shown in Figure 4-43, the (lack of) differences between rooms do not change significantly, and sleepiness under TDDs is still not significantly different than under electric lighting (dif. = -0.17, t(104) = 0.59, ns).

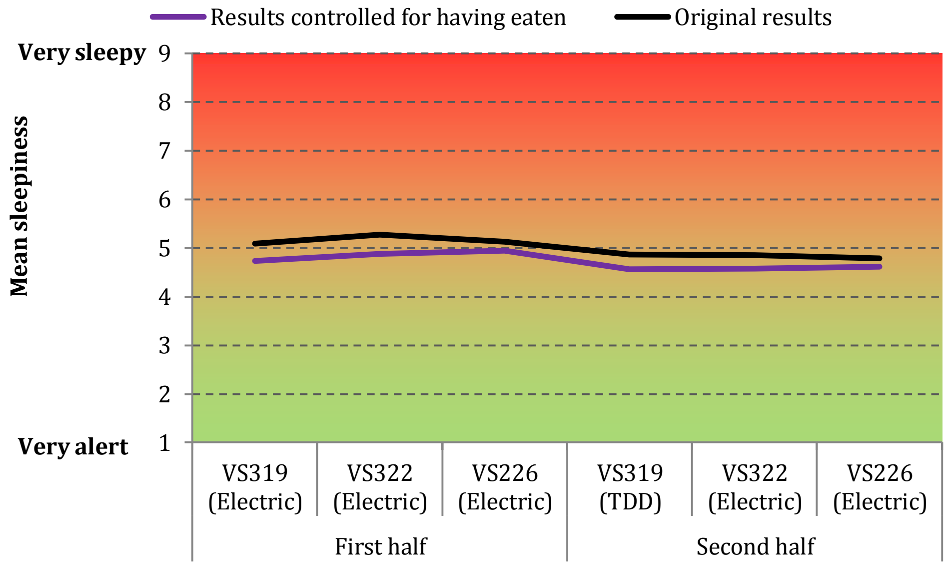


Figure 4-43: Sleepiness results controlled for having eaten.

The conclusions drawn from the change in arousal results also did not change when subjects having eaten was controlled for. As shown below, the overall pattern of results across the rooms did not change significantly, and differences between the rooms were still not significant.

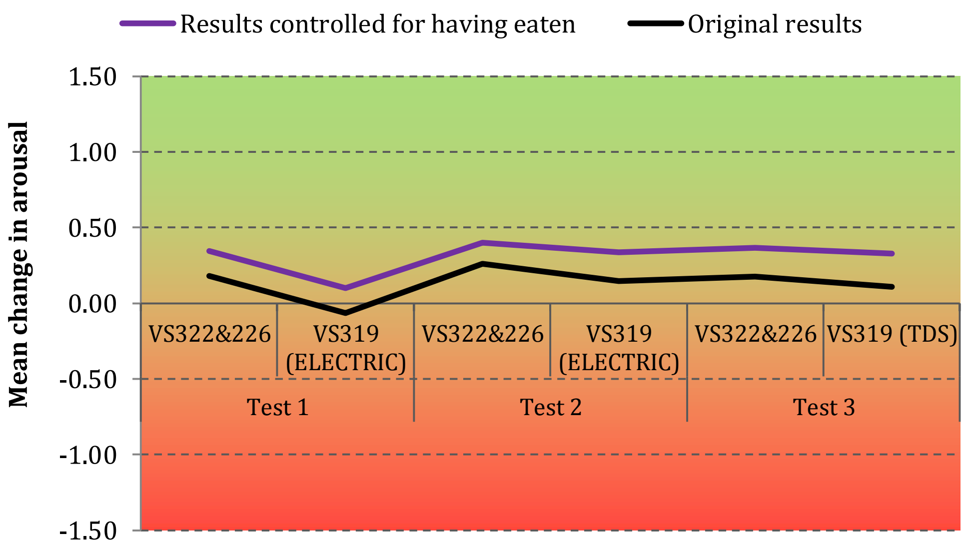


Figure 4-44: Change in arousal results controlled for having eaten.

4.2.2.2.2 Coffee drinking

Comparisons between those who had drunk coffee to those who hadn't found significant differences with regards to ratings of **illumination** (dif. = 0.64, $t(85) = 2.22$, $p<0.05$), while differences in ratings of **room attractiveness** (dif. = 0.43, $t(85) = 1.42$, ns) approached significance, with people who had drunk coffee perceiving the room as dimmer, and tending to regard the room as less attractive. There were also significant differences in **working memory** scores (dif. = 0.61, $t(84) = 2.22$, $p<0.05$) and **sleepiness** (dif. = 0.86, t(84) = 2.10, p<0.05), with those who drank coffee being less sleepy and having better performance. No significant differences were found for mood or processing speed.

However, controlling for the effects of drinking coffee does not change the results for perception of the room and lighting. The TDDs are still significantly better — both on average (Figure 4-45) and in the change in individual responses (Table 4-25).

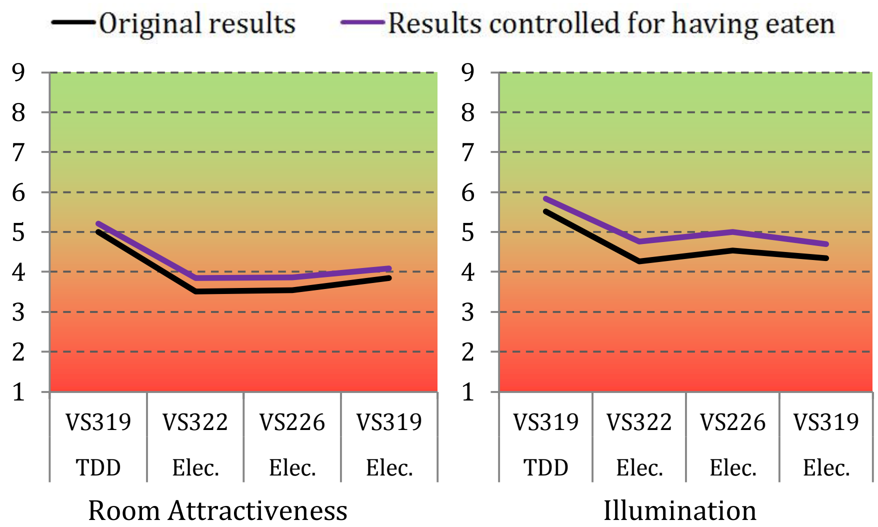


Figure 4-45: Perception results controlled for drinking coffee.

	Mean difference	SD	t(24)	significance	Original significance
Room Attractiveness	1.30	1.98	3.29	p<0.005	p<0.005
Illumination	1.04	1.55	3.35	p<0.005	p<0.005

Table 4-25: TDDs vs. Electric lighting in VS319: Paired comparison using the differences in individuals' responses in each situation with responses controlled for having drunk coffee.

Controlling for drinking coffee also does not change the results for performance: comparisons between the groups still find no significant effects (Figure 4-46, Table 4-26). This is not surprising, as the proportion of coffee drinkers in the groups does not vary that much — ranging from 15-33%.

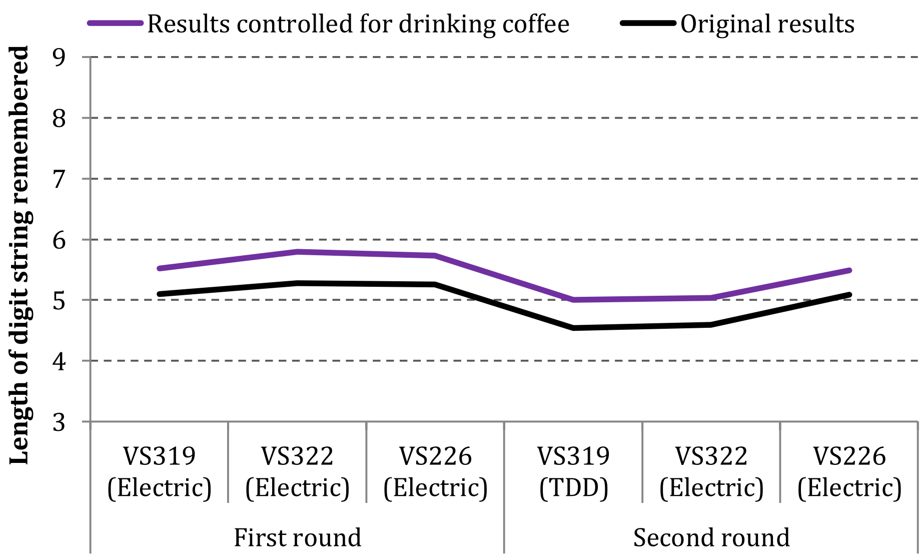


Figure 4-46: Digit Span results controlled for drinking coffee.

	n	Mean difference	SD	t	significance	Original significance
VS319	18	-0.02	0.89	-0.10	ns	ns
VS322 & 226	18	-0.06	1.58	-0.15	ns	ns

Table 4-26: Working memory: Round 1 vs. Round 2: Paired comparison using the differences in individuals' results in the two rounds of tests with results controlled for drinking coffee.

Neither does controlling for coffee drinking change the results for sleepiness: differences between the rooms are still not significant (Figure 4-47), and sleepiness under TDDs is not different to what it is under electric lighting (dif. = -0.27, t(104) = 0.92, ns).

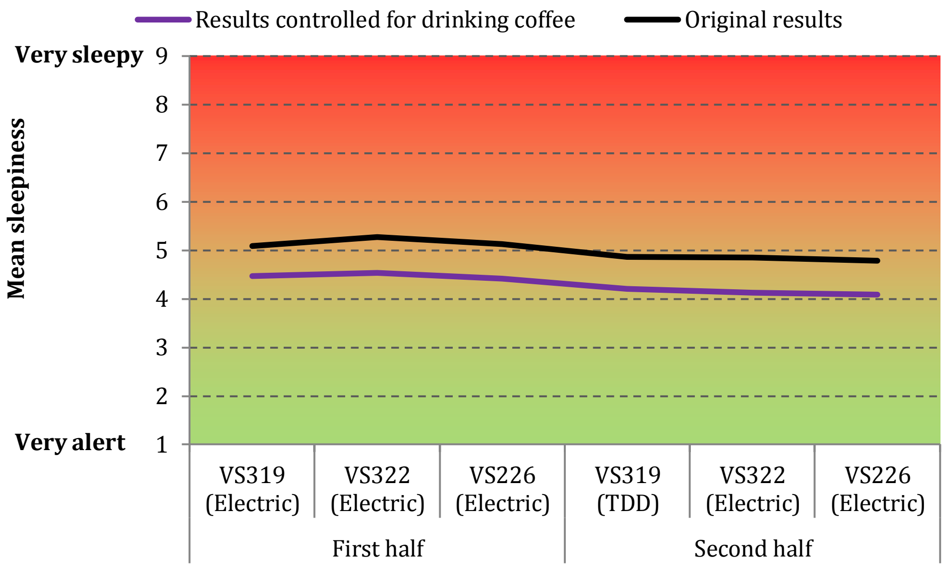


Figure 4-47: Sleepiness results controlled for drinking coffee.

4.2.2.2.3 Sleepiness

Sleepiness differs from the other controls in that it is a scale, rather than a binary option. Thus, the correlation between reported sleepiness and subject's responses was examined instead of simply breaking the sample into two groups.

Additionally, the correlation between people's *change* in response and their change in sleepiness between tests can be measured. This helps deal with the confounding effects of individual variation, enabling closer examination of the possible effects of sleepiness on the results. This comparison wasn't done for the "coffee drinking" and "having eaten" factors, because very few people changed their status between tests.

Perception

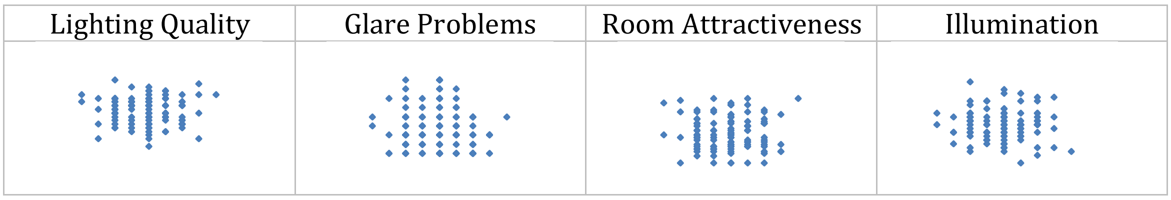


Figure 4-48: Scatter plots of perception scores (y-axis) against sleepiness (x-axis)

The above scatter plots show no correlations between sleepiness and subjects' responses to the lighting and rooms. This may be due to it being hidden by the confounding effects of the differences between individuals.

If the *changes* in individuals' responses between TDDs and electric lighting are compared to their *change* in sleepiness, then there are some significant correlations (Table 4-27) — suggesting that changes in sleepiness could be influencing people's responses.

Lighting quality	Glare problems	Room attractiveness	Illuminance
-0.29	-0.15	-0.48	-0.62

Table 4-27: Correlations between change in perception from electric lighting to TDDs and the subject's change in sleepiness. Note: significant ($p<0.05$) correlations are highlighted green, while non-significant ones are grey.

This is controlled for in the paired comparisons by adjusting the average change in response by the regression coefficient multiplied by the average change in sleepiness. The adjusted results are reported below:

	Mean difference	SD	t(24)	significance	Original significance
Room Attractiveness	1.29	1.95	3.31	p<0.005	p<0.005
Illumination	1.01	1.48	3.41	p<0.005	p<0.005

Table 4-28: TDDs vs. Electric lighting in VS319: Paired comparison of the differences in individuals' responses in each situation, controlling for the change in sleepiness.

However as can be seen, controlling for the change in sleepiness changes nothing — the TDDs still get significantly better responses than the electric lighting (Table 4-28). This is because the average change in sleepiness is very little.

Performance

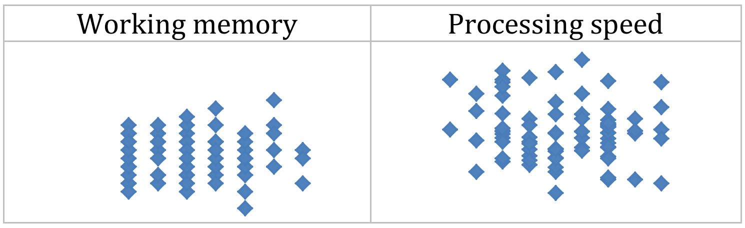


Figure 4-49: Scatter plots of performance scores (y-axis) against sleepiness (x-axis) under electric lighting

As with perception, there is no simple correlation between sleepiness and performance scores (Figure 4-49).

If the individuals' changes in performance are compared to their change in sleepiness — thus controlling for individual variation — then some possible correlations emerge:

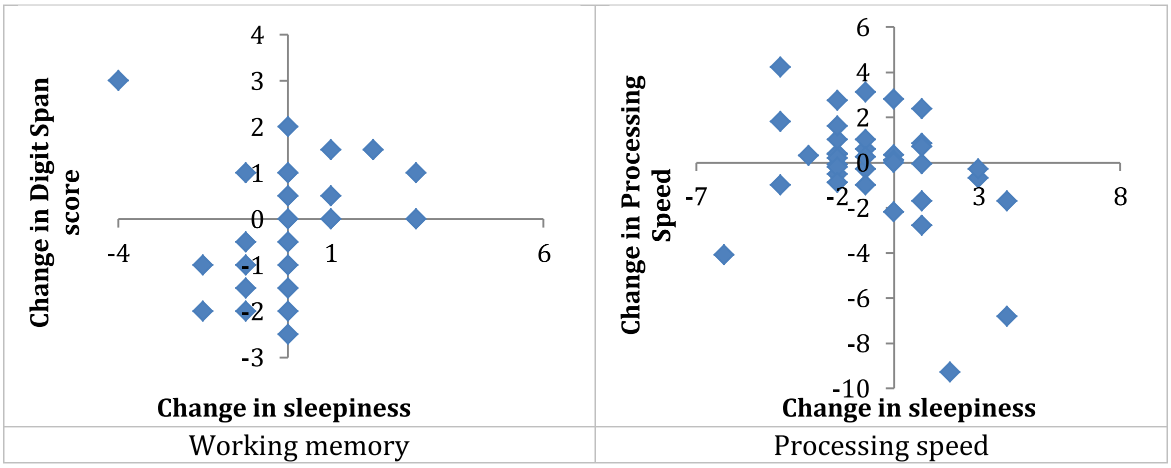


Figure 4-50: Scatter plots comparing subject's change in reported sleepiness to the change in their performance scores between the first and second rounds of testing

	Correlation	significance
Working memory	0.15	ns
... With possible outlier removed	0.45	p<0.01
Processing speed	-0.35	p<0.05

Table 4-29: Correlation of the change in performance scores to change in sleepiness between the two rounds of testing

Figure 4-50 indicates a *possible* effect of sleepiness on scores on the **working memory** test. The subjects that were *sleepier* the second time around got *higher test scores* than they did previously, while those that were *less sleepy* the second time around tended to perform worse. As shown in Table 4-29, the correlation *may* be significant — but only if a possible outlier is removed (in the top left corner in Figure 4-50). Such a relationship is somewhat counterintuitive and questionable — you would not expect people to do better on a cognitive performance test when they are *more* tired.

A more reasonable correlation is seen for **processing speed**, where lower sleepiness is correlated with better performance.

Regardless, both of these correlations suggest possible influencers, and so their effect on change in performance is controlled for to determine if they are an issue:

	n	Mean difference	SD	t	significance	Original significance
VS319	18	0.09	1.44	0.27	ns	ns
VS322 & 226	18	-0.10	1.60	-0.28	ns	ns

Table 4-30: Working memory: Round 1 vs. Round 2: Paired comparison using the differences in ratings from individual subjects in the two rounds of tests with results controlled for change in sleepiness.

	n	Mean difference	SD	t	significance	Original significance
VS319	21	-0.19	2.44	-0.36	ns	ns
VS322 & 226	20	0.26	3.18	0.37	ns	ns

Table 4-31: Processing Speed: Round 1 vs. Round 2: Paired comparison using the differences in ratings from individual subjects in the two rounds of tests with results controlled for change in sleepiness.

As seen in these tables, however, controlling for the change in sleepiness changes little — there are still no significant changes in performance between the rounds (Table 4-30,Table 4-31).

4.2.3 Overall: does controlling for these factors affect the results?

To sum up:

- Variation in air temperature, CO_2 levels, and noise levels does not appear to be significantly impacting the results.
- Perception results showed possible relationships to food intake, coffee consumption, and sleepiness. However, controlling for these factors did not change the comparisons between groups.
- Working memory results showed possible effects of coffee consumption, sleepiness, and needing glasses. Controlling for these factors did not change the comparisons between groups.
- Processing speed results showed possible effects of gender and sleepiness. Controlling for these factors did not change the comparisons between groups.
- Mood results only showed one possible effect, that of having eaten on change in arousal. Controlling for it did not change the results, and even if it had, it wouldn't have changed the factors that made the mood tests inconclusive.
- Sleepiness results showed effects of food intake and coffee consumption. However, controlling for these factors did not change the comparisons between groups.

In general, controlling for the different factors had very little effect on the results. This is because the variation between groups was generally small. For example, there is an average difference of 0.7 in scores of room attractiveness depending on whether or not people have eaten. However, the proportion of "people that have eaten" in a group only differs by about 15 percentage points — so the actual change to the results — as in the change in the *difference between groups,* is only about 0.1 — which isn't enough to change the results. Much of the time, the effects of controlling for the factors are even smaller than that.

The various factors examined here do not change the effects that have been found. It is thus possible to be reasonably confident that the changes in response observed under TDDs are due to the change in lighting, and not changes in these factors. In addition, none of these factors appear to have been covering up any other effects of TDDs.

4.3 What lighting factors can explain the effects?

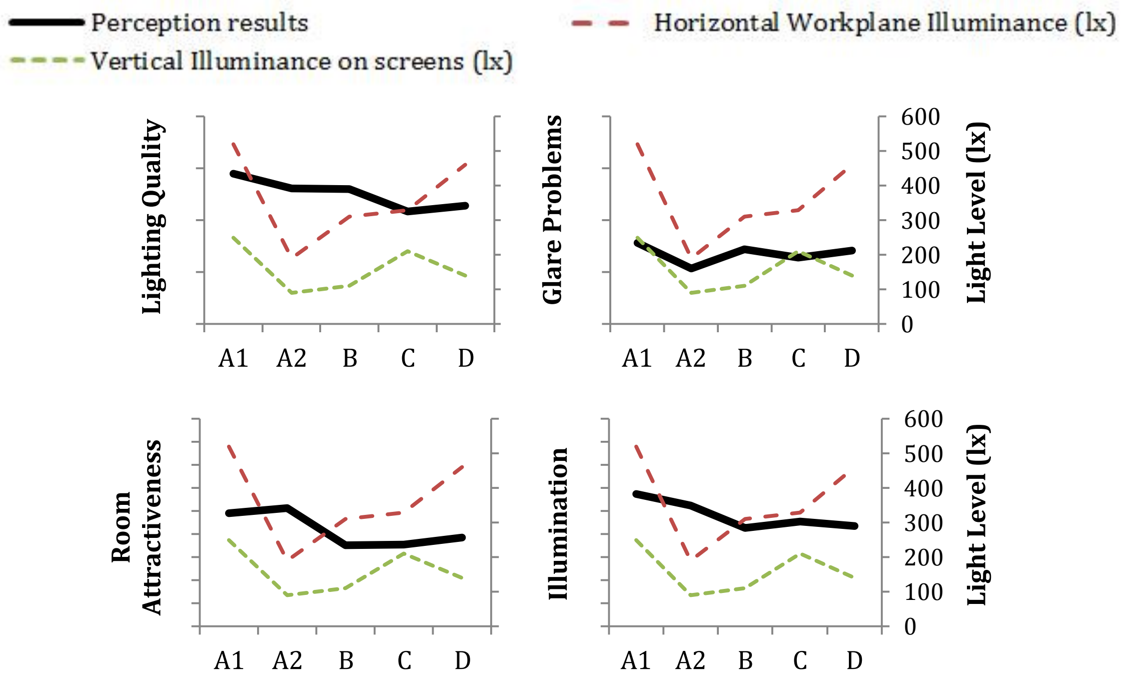
Several effects have been found that may be attributed to differences in the lighting, including the TDDs. After examining the possible effects of other environmental and demographic factors, it has been concluded that they have not significantly influenced the results. It appears then, that the lighting is probably the main cause of the effects.

The question now is *how* the lighting is affecting the results - what aspects of the lighting can be used to explain the effects?

This is important because the mechanisms behind the observed effects of the TDDs could affect the conclusions. For example, if the TDDs made the room more attractive merely because it was brighter, it would suggest that they are *not* inherently better than electric lighting, and that electric lighting could achieve the same results if it was brighter. If, however, the TDDs were better because the spectrum of daylight is better, then it would suggest that it has certain properties that make it inherently better than electric light.

Several possible explanations for the observed effects are discussed below. They are: light level, light distribution, spectrum, luminaire design, and psychological effects such as desire for daylight.

4.3.1 Light level



A1: VS319: TDD (sunny) | **A2**: VS319: TDD (overcast) | **B**: VS322: Electric
C: VS226: Electric | **D**: VS319: Electric

Figure 4-51: Perceptions of the rooms and lighting compared to the measured light levels..

The simple comparison of peoples' perceptions of the rooms and lighting to the measured light levels shows that the light levels do not explain the results (Figure 4-51). The differences simply do not match.

This may be because of limitations in the way the light levels were measured — measures such as horizontal workplane illuminance may not adequately describe how people actually perceive the brightness of light in a space. That such measures are flawed has been suggested before (e.g. Cuttle, 2010), and they may fail when comparing different lighting designs because they poorly address differences in aspects such as distribution.

Between the sunny and overcast days, the TDDs provide a comparison of two distinctly different light levels using the *same lighting*, which allows a more controlled assessment of the effects of different light levels.

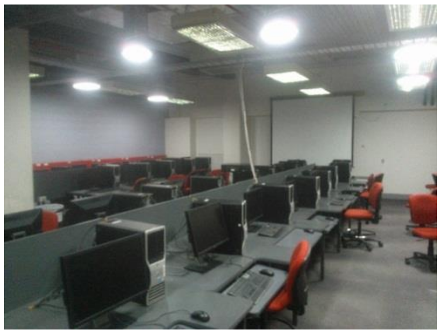
Comparing light levels in VS319 on both the sunny and the overcast day shows that the light levels were roughly 2.5 times as high on the sunny day. Despite this, as shown in section 4.1.1, the ratings of lighting quality and room attractiveness were the same. Thus, the improvements in VS319 from the TDDs are not (for the most part) due to changes in light level.

Light level does have some effects — people do perceive the room as being brighter on the sunny day, and glare problems may be more likely when it is brighter. However, the ratings of room attractiveness were not significantly different between the two days, and even on the overcast day, despite having significantly lower light levels than it did under electric lighting (at least as measured on the desktop), the room with TDDs was perceived as being "brighter".

Glare problems were not significantly different between the TDDs and electric lighting, however they *were* different between the sunny and overcast days, indicating that (unsurprisingly) light level does affect glare. However, the differences between the electric lighting and the TDDs are not large enough to be significant.

4.3.2 Light distribution

VS319 (Electric lighting)



VS319 (under TDDs)

Figure 4-52: Photos of VS319 with electric lighting and with TDDs

4.3.2.1 May be affecting room attractiveness

The change in light distribution is one of the most visible aspects of the TDDs. Where the electric lighting in all the rooms is focussed downwards, and casts noticeable shadows halfway down the walls, the TDD's light is much more diffuse (Figure 4-52). Studies suggest that high wall brightness is preferred by people (Flynn, 1977; Tenner, 2003). This may explain the improvement in room attractiveness, and the greater perceived brightness.

4.3.2.2 But not lighting quality

However, it is difficult to connect distribution to the change in perceived lighting quality. Subjects in VS322 report that it has significantly better lighting quality than the other rooms' electric lighting, but its light distribution is essentially the same, with the same focus on the horizontal plane and sharp shadows on the walls. Despite the fact that the TDDs provide very different lighting distribution to the lighting in VS322, their lighting quality is the same.

Similar findings have been found in other studies, where similar changes in lighting distribution between direct and indirect lighting also did not affect the reported lighting quality (Boyce et al., 2006a).

4.3.3 Spectrum

4.3.3.1 May also be affecting room attractiveness

The spectral characteristics or colour of the light from the TDDs is also distinctly different from the electric lighting. This could also explain the increase in room attractiveness under TDDs. One possible reason is the superior colour rendition of the daylight spectrum (Boyce et al., 2003), as studies of artificial lighting have found that lamps with higher colour rendering indices are preferred over lower ones, and are rated as giving better appearance (Veitch & McColl, 2001).

Similarly, spectrum differences may also partially explain the increase in perceived brightness, as studies have found that higher colour temperatures are perceived as brighter (Fotios & Levermore, 1997).

4.3.3.2 But not lighting quality

Like distribution, however, differences in light colour are a poor explanation for the differences in lighting quality, as the lamps in VS322 are the same colour as the rest of the electric lighting, yet its lighting quality is significantly different. Of course, it cannot be ruled out — it is possible that the light colour is enhancing the perceived lighting quality of the TDDs, while some other factor is enhancing it for the lighting in VS322. However, as discussed below, there is another, simpler explanation for the differences in lighting quality.

4.3.4 Luminaire design

Figure 4-53: Luminaire in VS226. Note the "egg-crate" style diffuser, and the lip around the fitting that suggests it should be recessed.

Figure 4-54: Luminaire in VS322.

Another possibility is that the luminaire design itself may affect people's perception of the lighting quality. In other words, the electric lighting in VS319 and VS226 (Figure 4-53), with its 90's era "egg-crate" style diffusers and suspended fittings that look like they should be recessed, may simply be perceived as being 'worse' than the louvered lighting in VS322 (Figure 4-54) and the TDDs. This dislike may lower the ratings of lighting quality. This explanation works well to explain why the TDDs and the lighting in VS322 are seen as providing better lighting quality. Other possible explanations involving light level, colour, or distribution, have the problem that the TDD's light is very visibly different from the electric lighting in all three rooms, but that VS322's light is not noticeably different from the electric lighting in the other rooms, as has already been discussed (see sections 4.3.1, 4.3.2, 4.3.3).

It is, of course, possible that higher Lighting Quality ratings for the electric lighting in VS322 are because Building Science students are more positive about lighting quality than Architecture or Interior Architecture students, or because of some other systematic difference between the groups. However, the explanation here is considered to be more likely as it is actually related to the lighting.

4.3.5 Psychological associations of daylight

Another possible explanation for the improvements in attractiveness and quality under TDDs has to do with demand effects. People might respond well to the TDDs because they *believe* daylight is better, or because they dislike fluorescent lights. Indications of such effects may be found by calculating the correlations between people's attitudes towards lighting and how their responses changed under TDDs.

	Lighting Quality	Glare	Room Attractiveness	Illumination
Sunny days make me happy	0.12	-0.17	0.04	0.38
Natural daylight indoors improves my mood	0.09	-0.12	0.01	0.19
Lack of sunlight in winter does not bother me	0.06	-0.17	0.17	-0.09
I do my best work in places that are lit using natural daylight	-0.24	-0.01	-0.06	-0.14
I get eyestrain from working under fluorescent lights	0.43	-0.27	0.50	0.37
Bright, harsh fluorescent lighting can make me feel tense	0.49	-0.15	0.40	0.32
It makes no difference to me what kind of lighting is in a room	-0.48	0.16	-0.39	-0.43

Table 4-32: Correlations between attitudes towards lighting and subject's change in perception of lighting under TDDs. Note: significant ($p<0.05$) correlations are highlighted green, while non-significant ones are grey. Marginal ($p<0.01$) correlations are highlighted blue. A full table of the correlations with all of the different questions in the lighting beliefs survey can be found in Appendix F5.

As shown in Table 4-32, people's attitude towards natural light does not seem to have much relation to why they like the TDDs lit room. There are, however, significant correlations between negative responses to fluorescent lights and the change in perception. There is also a negative correlation with not caring about the lighting in the room. This suggests that people who respond negatively to fluorescent lighting tend to appreciate the TDDs more, while people who don't really care about the lighting appreciate them less. This relationship is in line with that proposed by Cuttle (2002, in Boyce et al., 2003), where he suggested on the basis of survey data, that people's appreciation for daylight could be due to a belief in the negative effects of electric lighting.

A dislike of fluorescent lighting mediating the impact of TDDs is, in some ways, a positive feature, as "not-being-fluorescent-lighting" is not a quality that fluorescent lighting can replicate. On the other hand, a dislike of fluorescent lighting is not a quality of TDDs, but an attitude of some people. Moreover, attitudes can change, which makes such effects unreliable.

4.3.6 Summary

The main aspects of the TDDs that could be improving room attractiveness are most likely the distribution of the light, and its colour. The observed differences in perceptions of Lighting quality are more probably linked to the luminaire design, with the older luminaires in VS3319 and VS226 perhaps being seen as of lower quality.

It is also possible that the results are being influenced by a dislike of fluorescent lighting.

5 DISCUSSION AND CONCLUSIONS

5.1 Discussion: Are Tubular Daylighting Devices better than electric lighting in offices and educational facilities?

5.1.1 The electric lighting assessed can be considered 'typical'

The TDDs have been compared to the electric lighting in the computer rooms at the School of Architecture. To be able to extrapolate the findings out into the broader world, it is necessary to determine how representative the electric lighting in this study is of the typical lighting in offices and schools.

This was discussed in the methodology (3.2.1.3), and the key points are summarised below:

- The lighting layout is fairly normal.
- 4000K lamps are the typical ones used.
- The lighting in VS319 and VS226 is of a type common in the 1990's.
- The louver fittings in VS322 began to be used in the late 1980's, and were the most common fitting until recently. They are still used today because they are efficient. They may be considered to be typical post-1980's lighting.
- These days, designers try to use more indirect lighting, and get more diffuse light on the upper walls and ceilings and to reduce shadowing.

5.1.2 TDDs generally have comparable lighting quality to electric lighting

The TDDs showed similar lighting quality to the lighting in VS322, which, through the 1990's until recently, were the most common kind of electric lighting installed in offices, and are still used today. They have better lighting quality than the 90's-era lighting in VS319 and VS226.

The differences may be linked to the perceived quality of the design of the luminaires. This would suggest that to improve lighting quality, one should try to redesign the fittings to look good. This is possible for both TDDs and electric lighting, although electric lighting might have more flexibility.

The fact that the lighting quality does not seem to be affected significantly by changes in the light distribution suggests that the use of more indirect and diffuse light in good contemporary lighting might not have enhanced its lighting quality beyond that of TDDs. This is supported by comparisons with ratings of lighting quality found in other studies: the studies by Boyce et al. (2006a), done in 2003, included assessments of both basic direct lighting similar to the lighting in VS322, as well as direct/indirect lighting with an uplight component that is considered to be good practice in contemporary lighting. Both designs had lighting quality of ~2.8, which is about the same as the TDDs here. This suggests that the TDDs will compare well to contemporary lighting in terms of lighting quality.

If any modern lighting does have higher lighting quality, then the gap may be able to be overcome by designing a "nicer looking" appearance for the TDDs.

To conclude: TDDs provide comparable lighting quality to typical post-1980's electric lighting, which means it can be considered to be at least comparable to a significant amount of existing lighting in offices and schools — and it works just as well on both sunny and overcast days. Evidence also suggests that it is comparable to good contemporary lighting.

5.1.3 TDDs do not have any more glare than most electric lighting

The results show that the TDDs have "low" level glare problems, equal to electric lighting. Analysis indicates that glare problems are at least partly related to light level, with sunny days being slightly more glare prone than overcast days. The glare from the TDDs, however, was no different than it was from the electric lighting — even on sunny days.

Comparisons with the studies by Boyce et al. (2006a) indicate that the levels of glare in this study were normal, but that good modern indirect/direct lighting may be slightly better. However, as glare was low for all of the lighting in this study, it is not felt to be a major concern.

5.1.4 TDDs may make rooms more attractive than electric lighting

The TDDs can make rooms appear more attractive than typical post-1980's electric lighting. They also make it appear brighter, without using more power. Possible mechanisms for this effect include: the change in light distribution, with more light being spread out to the walls; the different spectrum of daylight; individual dislike of fluorescent lighting.

Distribution is not an advantage for TDDs

Light distribution, as a mechanism, is not a great advantage for TDDs. While it may allow it to be better than older lighting, electric lighting is perfectly capable of providing the same kind of light distribution as TDDs. More to the point, the change in distribution is similar to that provided by good contemporary lighting — i.e. more indirect and wall lighting. This mechanism, therefore, is not expected to advantage TDDs over good contemporary lighting, though it may be comparable.

Spectrum *can* be an advantage for TDDs

If the spectrum of the light from TDDs helps make the room more attractive, then that gives it an advantage over typical modern lighting, as 4000K lamps like the ones in this study are still normal practice (Hirschberg, 2012).

However, it is also possible that electric lighting may be able to get the same kind of effects with lamps of a higher colour temperature — for example, 17000K lamps have been found to positively affect people compared to 4000K ones in several studies (Mills et al., 2007; Rautkylä et al., 2010; Viola et al., 2008) — though they have not yet been compared against daylight.

Negative responses to fluorescents would be a solid advantage of TDDs — for now

People having a negative response to fluorescents is a potential mechanism for making TDDs more attractive than electric lighting. This factor could make TDDs better than *all* fluorescent lighting in offices and schools — over half the people in this study expressed negative responses to fluorescents.

However, it is possible that in the future LED lighting would also be able to replicate the feat of not being fluorescent lighting.

To conclude: TDDs can make rooms significantly more attractive than typical post-1980's electric lighting. There is also reason to hypothesise that it would be comparable to, and even slightly better than, good contemporary lighting — although without the advantage of distribution that it has over lighting in this study, its benefits would likely be lessened.

5.2 Conclusions

This study investigated the question of whether or not Tubular Daylighting Devices have better effects on people than electric lighting in offices and schools.

The literature on TDDs is sparse, and provides little information. The broader literature on daylighting suggests that daylighting through windows is preferred to electric lighting, and if done well makes spaces more appealing. It also suggests that daylighting can promote good health in offices and hospitals. Daylighting may also help improve people's mood, enhance alertness and reduce fatigue and stress — although effects may interact with a large range of factors, and may not be reliable. There is some evidence of possible performance and productivity effects, but studies in the area have had limited success, and overall the literature is inconclusive. The literature also shows that, when considering the benefits of windows, view is at least as important as daylight — if not more — and suggests that the high illuminances that windows deliver to the eye are also important. These factors suggest that TDDs cannot be as beneficial as windows, as they lack such properties.

In this study, TDDs were installed in a windowless computer room in the university, using a layout designed to be able to completely replace the electric lighting in the room. Tests were run on students in the room, under both electric lighting and the TDDs, as well as in two other computer rooms that were used as controls. This permitted the comparison of the students' responses to the different conditions, allowing us to identify how the TDDs affected their perception of the room and lighting, their mood, their sleepiness, and their performance.

So, do TDDs have better effects than electric lighting for schools and offices?

In some respects, they are better than typical post-1980's lighting. In no test did they perform worse than the electric lighting. They did make the room significantly more attractive, and was perceived as brighter. Despite being brighter, they did not have any more glare than the electric lighting. Its lighting quality was on a par with the most common form of post-1980's electric lighting, and was superior to the low quality 1990's lighting. Its lighting quality is likely comparable to good contemporary electric lighting, and its effect on room attractiveness may also be on a par with good contemporary lighting, if not slightly better — although further research would be needed to confirm this. Therefore, it would likely be an upgrade for many buildings.

Overall, it is suggested that TDDs provide a lighting environment at least as good as good quality contemporary lighting.

Some caution must be advised however. The scope of this study was limited to windowless computer rooms. Rooms that are significantly different, such as ones with windows, may not see the same kind of impact. Indeed, it is quite plausible that the most visible effects of the TDDs — the light colour and broadened distribution — would be less apparent in a space that already had natural light coming in the windows and providing vertical illumination.

Similarly, the sample itself was limited to university students in a school of architecture. It is possible that other groups may respond to lighting differently — studies have, for example, found effects of age on response to lighting (Knez & Kers, 2000).

The observed benefits are also limited. No direct performance or alertness effects were found, and its demonstrated advantages lie purely in making the room more attractive. Other studies (e.g. Veitch et al., 2011) have devised models suggesting links between room attractiveness and people's mood and performance, and it is quite possible that benefits from exposure to TDDs may be revealed over a longer term exposure typical of that in offices and schools. However, this study, with limited samples and shorter exposure, did not find evidence for such effects.

The issue of whether or not TDDs can be demonstrated to be a clearly beneficial and economic lighting option is still open. In some ways this research improves the case, as the demonstrated benefits to room attractiveness provided by the TDDs mean that the case for them can now be made through both the "daylight stimulates the circadian system" argument and the "people feel better and work better in more attractive rooms" argument. The lack of observed effects with regards to mood and performance does not necessarily count against them, as previous research has shown that finding such effects can be difficult and unreliable, and that such effects are more likely to appear in subjects exposed to the lighting for more prolonged periods.

5.3 Further study

This study leaves a number of unanswered questions worthy of further research.

First, there is still the question of whether or not TDDs can enhance productivity. Based on the lack of performance effects found in this study, and the response of Boyce et al. (2006b) to similar results in other lighting studies, we would recommend that further research in this area be field studies that study people exposed to the lighting for prolonged periods. Not only would this make it far more likely that effects (if they exist) would be strong enough to be observed, but it would also allow potential *health* effects to be examined — which would be more likely to find effects, as daylighting and health studies have been more consistently positive than the mood or performance studies. Moreover, the problems with assessing performance with short tests rather than people's productivity over extended periods remain an issue. Indirect measures, such as health, that can clearly be related to productivity are a good way to get around those problems. Some of the issues in this study also highlighted the difficulties in investigating mood. If mood is to be studied, it would be best if there was the opportunity to make many assessments over a prolonged period, so that the variation in mood can be averaged out and overall trends can be examined. Studies of different situations — schools, offices, hospitals — would all be good, as effects may vary between them.

Another question is how TDDs compare to good contemporary lighting. While educated inferences may be made from this study and others, it is necessary to actually test the hypotheses to confirm or deny them. If TDDs are to compete with good quality contemporary lighting, then this is important.

Thirdly, we may ask how TDDs interact with windows. There are two points of interest here. First, there is the question of how TDDs *compare* to windows. While it may be reasonably argued that windows would be better responded to because of the view they provide, it would be useful to know how significant the difference is. Similarly, it would be useful to compare TDDs and skylights. The other issue is how windows and TDDs interact. TDDs may have a significant impact on a room's lighting if there are no windows — but what about when there are? Does replacing the electric lighting with TDDs have the same kind of impact when there is already daylighting present? As has already been discussed, it is not an unreasonable hypothesis that the presence of windows would reduce the magnitude of any benefits derived from replacing electric lighting with TDDs. The change in light colour would likely be less noticeable, and side-lighting from windows could make the broadened light distribution of the TDDs less visible. If TDDs are to have broad application, then this is an important question to address.

Finally, there is the need for a more focussed investigation of the mechanisms by which TDDs may affect people's perceptions of the lighting and the room they're in. This study suggested several possible lighting-related factors that could be the reason for people's differing responses towards the different lighting designs. Studies to break down and assess the effects of each factor individually could be very useful not only to improve the design of TDDs, but also to improve lighting design in general.

This study has shown how Tubular Daylighting Devices compare to typical electric lighting from a human perspective. By doing so, it has expanded our knowledge of how people respond to TDDs, and has provided a basis for further research.

~finis~

6 REFERENCES

Akerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *The International journal of neuroscience*, *52*(1-2), 29–37.

Aries, M. (2005). *Human Lighting Demands - Healthy Lighting in an Office Environment*. Technische Universieit Eindhoven, Eindhoven. Retrieved from http://alexandria.tue.nl/extra2/200512454.pdf

Aries, M. B. C., Veitch, J. A., & Newsham, G. R. (2010). Windows, view, and office characteristics predict physical and psychological discomfort. *Journal of Environmental Psychology*, *30*(4), 533–541. doi:10.1016/j.jenvp.2009.12.004

Banerjee, P., Chatterjee, P., & Sinha, J. (2012). Is It Light or Dark? Recalling Moral Behavior Changes Perception of Brightness. *Psychological Science*, *23*(4), 407–409. doi:10.1177/0956797611432497

Banu, M. (2007). An experimental study on the appraisal of the visual environment at offices in relation to colour temperature and illuminance. *Building and Environment*, *42*(2), 979–983. doi:10.1016/j.buildenv.2005.10.022

Baron, R. A., Rea, M. S., & Daniels, S. G. (1992). Effects of indoor lighting (illuminance and spectral distribution) on the performance of cognitive tasks and interpersonal behaviors: The potential mediating role of positive affect. *Motivation and Emotion*, *16*(1), 1–33. doi:10.1007/BF00996485

Barrett, L. F., & Russell, J. A. (1999). The structure of current affect: Controversies and emerging consensus. *Current Directions in Psychological Science*, *8*(1), 10–14. doi:10.1111/1467-8721.00003

Beauchemin, K. M., & Hays, P. (1996). Sunny hospital rooms expedite recovery from severe and refractory depressions. *Journal of affective disorders*, *40*(1-2), 49–51.

Beauchemin, K. M., & Hays, P. (1998). Dying in the dark: sunshine, gender and outcomes in myocardial infarction. *Journal of the Royal Society of Medicine*, *91*(7), 352–354.

Begemann, S. H. A., van den Beld, G. J., & Tenner, A. D. (1997). Daylight, artificial light and people in an office environment, overview of visual and biological responses. *International Journal of Industrial Ergonomics*, *20*(3), 231–239. doi:10.1016/s0169-8141(96)00053-4

Boerstra, A., Beuker, T., Loomans, M., & Hensen, J. (2013). Impact of available and perceived control on comfort and health in European offices. *Architectural Science Review*, *0*(0), 1–12. doi:10.1080/00038628.2012.744298

Boray, P. F., Gifford, R., & Rosenblood, L. (1989). Effects of warm white, cool white and full-spectrum fluorescent lighting on simple cognitive performance, mood and ratings of others. *Journal of Environmental Psychology*, *9*(4), 297–307. doi:10.1016/S0272-4944(89)80011-8

Boubekri, M. (1995). Appraisal of the Lighting Conditions in an Office Building: Results of a Survey. *Indoor and Built Environment*, *4*(3-4), 162–169. doi:10.1177/1420326X9500400306

Boubekri, Mohamed, Hull, R. B., & Boyer, L. L. (1991). Impact of Window Size and Sunlight Penetration on Office Workers' Mood and Satisfaction. *Environment and Behavior*, *23*(4), 474–493. doi:10.1177/0013916591234004

Boyce, P. R. (1974). Illuminance, difficulty, complexity and visual performance. *Lighting Research and Technology*, *6*(4), 222–226. doi:10.1177/096032717400600406

Boyce, P. R. (2003). *Human Factors in Lighting* (2nd ed.). London ; New York: Taylor & Francis.

Boyce, P. R. (2004). *Reviews on Technical Reports on Daylight and Productivity*. Lighting Research Center, Troy, NY: Capturing the Daylight Dividends Program. Retrieved from http://www.daylighting.org/BoyceHMGReview.pdf

Boyce, P. R., & Cuttle, C. (1990). Effect of correlated colour temperature on the perception of interiors and colour discrimination performance. *Lighting Research and Technology*, *22*(1), 19–36. doi:10.1177/096032719002200102

Boyce, P. R., Hunter, C., & Howlett, O. (2003). *The Benefits of Daylight through Windows*. U.S. Department of Energy. Retrieved from http://www.lrc.rpi.edu/programs/daylighting/pdf/DaylightBenefits.pdf

Boyce, P. R., Veitch, J. A., Newsham, G. R., Jones, C. C., Heerwagen, J., Myer, M., & Hunter, C. M. (2006a). Lighting quality and office work: two field simulation experiments. *Lighting Research and Technology*, *38*(3), 191–223. doi:10.1191/1365782806lrt161oa

Boyce, P. R., Veitch, J. A., Newsham, G. R., Jones, C. C., Heerwagen, J., Myer, M., & Hunter, C. M. (2006b). Authors' response. *Lighting Research and Technology*, *38*(4), 377–378. doi:10.1177/147709380603800421

BRANZ Ltd. (2007). Designing Quality Learning Spaces: Ventilation and Indoor Air Quality. *Ministry of Education*. Retrieved 25 February 2013, from http://www.minedu.govt.nz/~/media/MinEdu/Files/EducationSectors/PrimarySecondary/PropertyToolbox/ModernLearning/VentilationIndoorAirQualityGuide.pdf

Buchanan, T. L., Barker, K. N., Gibson, J. T., Jiang, B. C., & Pearson, R. E. (1991). Illumination and errors in dispensing. *American journal of hospital pharmacy*, *48*(10), 2137–2145.

Çakir, A., & Çakir, G. (1998). *Light and Health*. Berlin: ERGONOMIC Institute. Retrieved from http://www.healthylight.de/Light_and_Health/Documents_files/1LightandHealth.pdf

Carter, D. J., & Al Marwaee, M. (2009). User attitudes toward tubular daylight guidance systems. *Lighting Research and Technology*, *41*(1), 71–88. doi:10.1177/1477153508096045

Charles, K. E., & Veitch, J. A. (2002). *Environmental Satisfaction in Open-Plan Environments: 2. Effects of Workstation Size, Partition Height and Windows*. Ottawa, ON: Institute for Research in Construction, National Research Council, Canada. Retrieved from http://www.nrc-cnrc.gc.ca/obj/irc/doc/pubs/ir/ir845/ir845.pdf

Clapham, M. M. (2004). The Convergent Validity of the Torrance Tests of Creative Thinking and Creativity Interest Inventories. *Educational and Psychological Measurement*, *64*(5), 828–841. doi:10.1177/0013164404263883

Cohen, A. (2011). The photosynthetic President: Converting sunshine into popularity. *The Social Science Journal*, *48*(2), 295–304. doi:10.1016/j.soscij.2010.11.007

Coley, D. A., & Greeves, R. (2004). *The effect of low ventilation rates on the cognitive function of a primary school class* (No. R102). Exeter University.

Collins, B. L. (1976). Review of the Psychological Reaction to Windows. *Lighting Research and Technology*, *8*(2), 80–88. doi:10.1177/14771535760080020601

Cunningham, M. R. (1979). Weather, Mood, and Helping Behavior: Quasi Experiments with the Sunshine Samaritan. *Journal of Personality and Social Psychology*, *37*(11), 1947–1956. doi:10.1037/0022-3514.37.11.1947

Cuttle, C. (2002). Identifying the human values associated with windows. *International Daylighting*, (5), 3–6.

Cuttle, C. (2010). Towards the third stage of the lighting profession. *Lighting Research and Technology*, *42*(1), 73–93. doi:10.1177/1477153509104013

Dasgupta, U. (2003). *The Impact of Windows on Mood and the Performance of Judgmental Tasks* (M.S in Lighting). Rensselaer Polytechnic Institute.

Demos, G. D., Davis, S., & Zuwaylif, F. F. (1967). Controlled physical environments. *Building Research*, *4*, 60–62.

Dow, G. (2003). Creativity Test: Torrance Tests of Creative Thinking (1962). Retrieved 30 April 2012, from http://www.indiana.edu/~bobweb/Handout/d3.ttct.htm

Edwards, L., & Torcellini, P. (2002). *A Literature Review of the Effects of Natural Light on Building Occupants*. Colorado: National Renewable Energy Laboratory. Retrieved from http://www.nrel.gov/docs/fy02osti/30769.pdf

Ejhed, J. (2001). Daylight quantities acceptance studies in the Arthelio project. In *9th Lux Europa Congress* (pp. 22–25). Reykjavik.

European Agency for Safety and Health at Work. (2005). *Noise in figures*. Belgium: Luxembourg: Office for Official Publications of the European Communities. Retrieved from https://osha.europa.eu/en/publications/reports/6905723

Figueiro, M. G., Brons, J. A., Plitnick, B., Donlan, B., Leslie, R. P., & Rea, M. S. (2011). Measuring circadian light and its impact on adolescents. *Lighting Research and Technology*, *43*(2), 201–215. doi:10.1177/1477153510382853

Finnegan, M. C., & Solomon, L. Z. (1981). Work attitudes in windowed vs. windowless environments. *Journal of Social Psychology*, *115*(2), 291.

Flynn, J. E. (1977). A study of subjective responses to low energy and nonuniform lighting systems. *Lighting Design And Application*, *7*(2).

Fostervold, K. I., & Nersveen, J. (2008). Proportions of direct and indirect indoor lighting — The effect on health, well-being and cognitive performance of office workers. *Lighting Research and Technology*, *40*(3), 175–200. doi:10.1177/1477153508090917

Fotios, S., & Gado, T. (2005). A comparison of visual objectives used in side-by-side matching tests. *Lighting Research and Technology*, *37*(2), 117–130. doi:10.1191/1365782805li133oa

Fotios, S., & Levermore, G. J. (1997). Perception of electric light sources of different colour properties. *Lighting Research and Technology*, *29*(3), 161–171. doi:10.1177/14771535970290030701

Galasiu, A. D., & Veitch, J. A. (2006). Occupant preferences and satisfaction with the luminous environment and control systems in daylit offices: a literature review. *Energy and Buildings*, *38*(7), 728–742.

Gifford, R. (1994). *Scientific Evidence for Claims about Full-Spectrum Lamps: Past and Future* (IRC Internal Report No. 659) (pp. 37–46). Ottawa, ON: National Research Council of Canada, Institute for Research in Construction. Retrieved from http://www.nrc-cnrc.gc.ca/obj/irc/doc/pubs/ir/ir659/gifford.pdf

Goncalo, J. A., Flynn, F. J., & Kim, S. H. (2010). Are Two Narcissists Better Than One? The Link Between Narcissism, Perceived Creativity, and Creative Performance. *Personality and Social Psychology Bulletin*, *36*(11), 1484–1495. doi:10.1177/0146167210385109

Gutkowski, J. M. (1992). *The Impact of Windows on Positive Affect and Cognitive Performance in Work Settings* (M.S.). Rensselaer Polytechnic Institute.

Harter, J. K., Schmidt, F. L., & Hayes, T. L. (2002). Business-Unit-Level Relationship Between Employee Satisfaction, Employee Engagement, and Business Outcomes: A

Meta-Analysis. *Journal of Applied Psychology*, *87*(2), 268–279. doi:10.1037/0021-9010.87.2.268

Hartleb, S. B. (1989). *Some Effects of the Sequential Experience of Windows on Human Response* (M.S.). Rensselaer Polytechnic Institute.

Hawes, B. K., Brunyé, T. T., Mahoney, C. R., Sullivan, J. M., & Aall, C. D. (2012). Effects of four workplace lighting technologies on perception, cognition and affective state. *International Journal of Industrial Ergonomics*, *42*(1), 122–128. doi:10.1016/j.ergon.2011.09.004

Hedge, A. (1994). Reactions of computer users to three different lighting systems in windowed and windowless offices. *Work and Display Units, '94*, B54–B56.

Heerwagen, J., & Heerwagen, D. (1986). Lighting and psychological comfort. *Lighting Design And Application*, *16*(4), 47–51.

Hendrick, C., Martyniuk, O., Spencer, T. J., & Flynn, J. E. (1977). Procedures for Investigating the Effect of Light on Impression. *Environment and Behavior*, *9*(4), 491–510. doi:10.1177/001391657794003

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2-3), 61–83. doi:10.1017/S0140525X0999152X

Heschong, L., Wright, R. L., & Okura, S. (2002). Daylighting Impacts on Human Performance in School. *Journal of the Illuminating Engineering Society*, *31*, 101–114.

Heschong Mahone Group. (1999a). *Daylighting in Schools: An investigation into the relationship between daylighting and human performance*. Sacramento. Retrieved from http://www.h-m-g.com/downloads/Daylighting/schoolc.pdf

Heschong Mahone Group. (1999b). *Skylighting and Retail Sales: An investigation into the relationship between daylighting and human performance. Condensed report.* Retrieved from http://www.h-m-g.com/downloads/Daylighting/retailc.pdf

Heschong Mahone Group. (2003a). *Windows and Classrooms: A study of student performance and the indoor environment*. California Energy Commision. Retrieved from http://www.h-m-g.com/downloads/Daylighting/A-7_Windows_Classrooms_2.4.10.pdf

Heschong Mahone Group. (2003b). *Windows and Offices: A study of office worker performance and the indoor environment*. California Energy Commision. Retrieved from http://www.h-m-g.com/downloads/Daylighting/A-9_Windows_Offices_2.6.10.pdf

Heschong Mahone Group. (2003c). *Daylight and Retail Sales*. California Energy Commision. Retrieved from http://www.h-m-g.com/downloads/Daylighting/A-5_Daylgt_Retail_2.3.7.pdf

Hidajat, R. R., Hidayat, J. R., McLay, J. L., Elder, M. J., Goode, D. H., & Pointon, R. C. (2004). A fast system for reporting the Farnsworth-Munsell 100-hue colour vision test. *Documenta ophthalmologica. Advances in ophthalmology*, *109*(2), 109–114.

Hirschberg, R. (2012).

Hoonhout, H. C. M. (Jettie), Knoop, M., & Vanpol, R. (2009). Colored Lighting in Offices the New Caffeine? Looking into Performance Effects of Colored Lighting. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *53*(8), 502–506. doi:10.1177/154193120905300804

Hygge, S., & Knez, I. (2001). EFFECTS OF NOISE, HEAT AND INDOOR LIGHTING ON COGNITIVE PERFORMANCE AND SELF-REPORTED AFFECT. *Journal of Environmental Psychology*, *21*(3), 291–299. doi:10.1006/jevp.2001.0222

Isen, A. M. (2001). An Influence of Positive Affect on Decision Making in Complex Situations: Theoretical Issues With Practical Implications. *Journal of Consumer Psychology*, *11*(2), 75–85. doi:10.1207/S15327663JCP1102_01

Isen, A. M., Daubman, K. A., & Nowicki, G. P. (1987). Positive Affect Facilitates Creative Problem Solving. *Journal of Personality and Social Psychology*, *52*(6), 1122–1131. doi:10.1037/0022-3514.52.6.1122

Jackson, Q. M. (2006). *Daylighting in schools : a New Zealand perspective*. Victoria University of Wellington. Retrieved from http://restrictedarchive.vuw.ac.nz//handle/123456789/6390

Joarder, A. R., & Price, A. D. F. (2012). Impact of daylight illumination on reducing patient length of stay in hospitals after coronary artery bypass graft surgery. *Lighting Research and Technology*. doi:10.1177/1477153512455940

Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., & Fukasawa, K. (2006). Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology*, *117*(7), 1574–1581. doi:10.1016/j.clinph.2006.03.011

Katzev, R. (1992). The Impact of Energy-Efficient Office Lighting Strategies on Employee Satisfaction and Productivity. *Environment and Behavior*, *24*(6), 759–778. doi:10.1177/0013916592246004

Keep, P., James, J., & Inman, M. (1980). Windows in the intensive therapy unit. *Anaesthesia*, *35*(3), 257–262. doi:10.1111/j.1365-2044.1980.tb05093.x

Kilic, D. K., & Hasirci, D. (2011). Daylighting Concepts for University Libraries and Their Influences on Users' Satisfaction. *The Journal of Academic Librarianship*, *37*(6), 471–479. doi:10.1016/j.acalib.2011.07.003

Kinnear, P. R., & Sahraie, A. (2002). New Farnsworth-Munsell 100 hue test norms of normal observers for each year of age 5-22 and for age decades 30-70. *The British journal of ophthalmology*, *86*(12), 1408–1411.

Knez, I. (1995). Effects of indoor lighting on mood and cognition. *Journal of Environmental Psychology*, *15*(1), 39–51. doi:10.1016/0272-4944(95)90013-6

Knez, I. (2001). EFFECTS OF COLOUR OF LIGHT ON NONVISUAL PSYCHOLOGICAL PROCESSES. *Journal of Environmental Psychology*, *21*(2), 201–208. doi:10.1006/jevp.2000.0198

Knez, I., & Enmarker, I. (1998). Effects of Office Lighting on Mood and Cognitive Performance And A Gender Effect in Work-xRelated Judgment. *Environment and Behavior*, *30*(4), 553–567. doi:10.1177/001391659803000408

Knez, I., & Hygge, S. (2002). Irrelevant speech and indoor lighting: effects on cognitive performance and self-reported affect. *Applied Cognitive Psychology*, *16*(6), 709–718. doi:10.1002/acp.829

Knez, I., & Kers, C. (2000). Effects of Indoor Lighting, Gender, and Age on Mood and Cognitive Performance. *Environment and Behavior*, *32*(6), 817–831. doi:10.1177/0013916500326005

Kretschmer, V., Griefahn, B., & Schmidt, K.-H. (2011). Bright light and night work: effects on selective and divided attention in elderly persons. *Lighting Research and Technology*, *43*(4), 473–486. doi:10.1177/1477153511405808

Küller, R., & Wetterberg, L. (1993). Melatonin, cortisol, EEG, ECG and subjective comfort in healthy humans: Impact of two fluorescent lamp types at two light intensities. *Lighting Research and Technology*, *25*(2), 71–80. doi:10.1177/096032719302500203

Küller, Rikard, & Lindsten, C. (1992). Health and behavior of children in classrooms with and without windows. *Journal of Environmental Psychology*, *12*(4), 305–317. doi:10.1016/S0272-4944(05)80079-9

Larson, C. T., & others. (1965). The effect of windowless classrooms on elementary school children. Retrieved from http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED014847

Leaman, A., & Bordass, B. (2000). Productivity in buildings: The 'killer' variables. In D. Clements-Croome & F. N. Spon (Eds.), *Creating the Productive Workplace*. London.

Leather, P., Pyrgas, M., Beale, D., & Lawrence, C. (1998). Windows in the Workplace. *Environment and Behavior*, *30*(6), 739–762. doi:10.1177/001391659803000601

Lehrl, S., Gerstmeyer, K., Jacob, J. H., Frieling, H., Henkel, A. W., Meyrer, R., … Bleich, S. (2007). Blue light improves cognitive performance. *Journal of Neural Transmission (Vienna, Austria: 1996)*, *114*(4), 457–460. doi:10.1007/s00702-006-0621-4

Leyten, J. L., Kurvers, S. R., & Raue, A. K. (2012). Temperature, thermal sensation and workers' performance in air-conditioned and free-running environments. *Architectural Science Review*, *0*(0), 1–8. doi:10.1080/00038628.2012.745391

Marwaee, M. A., & Carter, D. J. (2006). A field study of tubular daylight guidance installations. *Lighting Research and Technology*, *38*(3), 241–258. doi:10.1191/1365782806lrt170oa

Mayhoub, M. S., & Carter, D. J. (2011). The costs and benefits of using daylight guidance to light office buildings. *Building and Environment*, *46*(3), 698–710. doi:10.1016/j.buildenv.2010.09.014

McCloughan, C. L. B., Aspinall, P. A., & Webb, R. S. (1999). The impact of lighting on mood. *Lighting Research and Technology*, *31*(3), 81–88. doi:10.1177/096032719903100302

Mehrabian, A. (1974). *An approach to environmental psychology*. Cambridge: M.I.T. Press.

Mills, P. R., Tomkins, S. C., & Schlangen, L. J. (2007). The effect of high correlated colour temperature office lighting on employee wellbeing and work performance. *Journal of Circadian Rhythms*, *5*(1), 2. doi:10.1186/1740-3391-5-2

Miner, A. G., & Glomb, T. M. (2010). State mood, task performance, and behavior at work: A within-persons approach. *Organizational Behavior and Human Decision Processes*, *112*(1), 43–57. doi:10.1016/j.obhdp.2009.11.009

Ministry of Economic Development. (2011). Prices | Ministry of Economic Development. Retrieved 3 March 2013, from http://www.med.govt.nz/sectors-industries/energy/energy-modelling/data/prices

Miwa, Y., & Hanyu, K. (2006). The Effects of Interior Design on Communication and Impressions of a Counselor in a Counseling Room. *Environment and Behavior*, *38*(4), 484–502. doi:10.1177/0013916505280084

Moore, D. S. (2000). *The Basic Practice of Statistics* (2nd ed.). USA: W H. Freeman and Company.

Moore, E. O. (1981). A Prison Environment's Effect on Health Care Service Demands. *Journal of Environmental Systems*, *11*(1), 1–1. doi:10.2190/KM50-WH2K-K2D1-DM69

Moore, T., Carter, D. J., & Slater, A. I. (2004). A study of opinion in offices with and without user controlled lighting. *Lighting Research and Technology*, *36*(2), 131–144. doi:10.1191/1365782804li109oa

Narvaez, D., & Bock, T. (2002). Moral Schemas and Tacit Judgement or How the Defining Issues Test is Supported by Cognitive Science. *Journal of Moral Education*, *31*(3), 297–314.

Newsham, G. R., Veitch, J. A., Arsenault, C. D., & Duval, C. L. (2004). Effect of dimming control on office worker satisfaction and performance. In *Proceedings of the IESNA Annual Conference*. Tampa, FL. Retrieved from http://www.nrc-cnrc.gc.ca/obj/irc/doc/pubs/nrcc47069/nrcc47069.pdf

Noguchi H, Itou T, Katayama S, Koyama E, Morita T, & Sato M. (2004). Effects of Bright Light Exposure in the Office. *J Physiol Anthropol Appl Hum Sci*, *23*(6), 368.

Oseland, N. (1999). *Environmental Factors Affecting Office Worker Performance: A Review of Evidence*. London: Chartered Institution of Building Services Engineers : DETR.

Pellegrino, A. (1999). Assessment of artificial fighting parameters in a visual comfort perspective. *Lighting Research and Technology*, *31*(3), 107–115. doi:10.1177/096032719903100305

Plant, C. G. . (1970). The Light of Day. *Light and Lighting*, 292–296.

Plitnick, B., Figueiro, M. G., Wood, B., & Rea, M. S. (2010). The effects of red and blue light on alertness and mood at night. *Lighting Research and Technology*, *42*(4), 449–458. doi:10.1177/1477153509360887

Rashid, M., & Zimring, C. (2008). A Review of the Empirical Literature on the Relationships Between Indoor Environment and Stress in Health Care and Office Settings. *Environment and Behavior*, *40*(2), 151–190. doi:10.1177/0013916507311550

Rautkylä, E., Puolakka, M., Tetri, E., & Halonen, L. (2010). Effects of Correlated Colour Temperature and Timing of Light Exposure on Daytime Alertness in Lecture Environments. *Journal of Light & Visual Environment*, *34*(2), 59–68.

Rest, J., & Narvaez, D. (1998). Defining Issues Test - 2. Univesity of Minnesota. Retrieved from http://www.liberalarts.wabash.edu/storage/assessment-instruments/dit2.pdf

Romm, J., & Browning, W. (1994). *Greening the building and the bottom line: Increasing productivity through energy-efficient design*. Rocky Mountain Institute. Retrieved from http://www.enterpriseelectricboise.com/documents/GreenBuildingsTheBottomLine.pdf

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, *11*(3), 273–294. doi:10.1016/0092-6566(77)90037-X

Russell, J. A., & Snodgrass, J. (1991). Emotion and the environment. In D. Stokols & I. Altman (Eds.), *Handbook of Environmental Psychology* (Vols. 1-2, Vol. 1, pp. 245–280). Malabar, Fla: Krieger Pub. Co.

Ruys, T. (1970). *Windowless Offices* (Masters Thesis). University of Washington.

Santamaria, J. G., & Bennett, C. A. (1980). Performance and Aesthetic Effects of Daylight. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *24*(1), 228–232. doi:10.1177/107118138002400162

Seppänen, O., Fisk, W. J., & Lei, Q. H. (2006). Effect of Temperature on Task Performance in Office Environment. In *Proceedings Cold Climate HVAC conference*. Moscow.

Shemirani, S. M. M., Memarian, G. H., Naseri, S. P., Nejad, H. H., & Vaziri, V. (2011). Investigating the Behaviors of the Elementary School Students in Reference to Factors Associated with Daylight. *Asian Social Science*, *7*(3), p237. doi:10.5539/ass.v7n3p237

Slater, A. I., Perry, M. J., & Carter, D. J. (1993). Illuminance differences between desks: Limits of acceptability. *Lighting Research and Technology*, *25*(2), 91–103. doi:10.1177/096032719302500208

Solatube. (2010). Solatube Daylighting System. *Solatube*. Retrieved 4 July 2013, from http://www.solatube.co.nz/downloads/product-literature/750ds-oz-new-2010_compressed.pdf

Solatube. (2012a). Solatube - What is Daylighting. Retrieved 3 March 2013, from http://www.solatube.co.nz/residential/what-is-daylighting/index.php

Solatube. (2012b). Solatube Commercial Daylighting Experience. Retrieved 25 February 2013, from http://www.solatube.co.nz/commercial/daylighting/commercial-daylighting-experience.php

Standards Association of Australia, & Standards New Zealand. (2006). *NZS 1680.1: Interior and workplace lighting*. Sydney, N.S.W. : Wellington, N.Z: Standards Australia ; Standards New Zealand.

Standards New Zealand. (1990). *NZS 4303: Ventilation for acceptable indoor air quality*. Wellington [N.Z.]: Standards New Zealand.

Standards New Zealand. (2007). *NZS 4243: Part 1: Energy efficiency: large buildings*. Wellington [N.Z.]: Standards New Zealand.

Standards New Zealand. (2008). *NZS 6801: Acoustics: measurement of environmental sound*. Wellington, [N.Z.]: Standards New Zealand.

Stone, N. J. (1998). Windows and Environmental Cues on Performance and Mood. *Environment and Behavior*, *30*(3), 306–321. doi:10.1177/001391659803000303

Stone, N. J., & Irvine, J. M. (1993). Performance, Mood, Satisfaction, and Task Type in Various Work Environments: A Preliminary Study. *The Journal of General Psychology*, *120*(4), 489–497. doi:10.1080/00221309.1993.9711162

Stone, N. J., & Irvine, J. M. (1994). Direct or indirect window access, task type, and performance. *Journal of Environmental Psychology*, *14*(1), 57–63. doi:10.1016/S0272-4944(05)80198-7

Sundstrom, E. D. (1986). *Work places: the psychology of the physical environment in offices and factories*. Cambridge [Cambridgeshire] ; New York: Cambridge University Press.

Tenner, A. D. (2003). A Healthy Future for Office Lighting? *Journal of Light & Visual Environment*, *27*(3), 172–176.

Thompson, J. (2012). *Are automated daylight control systems working as they should?* (Unpublished Masters Thesis). Victoria University of Wellington.

Ulrich, R. S. (1984). View through a window may influence recovery from surgery. *Science (New York, N.Y.)*, *224*(4647), 420–421.

Van Bommel, W. J. M. (2006). Non-visual biological effect of lighting and the practical meaning for lighting for work. *Applied Ergonomics*, *37*(4), 461–466. doi:10.1016/j.apergo.2006.04.009

Veitch, J A, & McColl, S. L. (2001). A critical examination of perceptual and cognitive effects attributed to full-spectrum fluorescent lighting. *Ergonomics*, *44*(3), 255–279. doi:10.1080/00140130121241

Veitch, J. A, & Newsham, G. R. (1998). Lighting quality and energy-efficiency effects on task performance, mood, health, satisfaction and comfort. *Journal of the Illuminating Engineering Society*, *27*(1), 107–129.

Veitch, J. A, & Newsham, G. R. (2000). Preferred luminous conditions in open-plan offices: research and practice recommendations. *Lighting Research and Technology*, *32*(4), 199–212. doi:10.1177/096032710003200404

Veitch, J. A., & Gifford, R. (1996). Assessing beliefs about lighting effects on health, performance, mood and social behaviour. *Environment and Behavior*, *28*(4), 446–470.

Veitch, J. A., Newsham, G. R., Boyce, P. R., & Jones, C. C. (2008). Lighting appraisal, well-being and performance in open-plan offices: A linked mechanisms approach. *Lighting Research and Technology*, *40*(2), 133–151. doi:10.1177/1477153507086279

Veitch, Jennifer A, & Newsham, G. R. (2000). Exercised Control, Lighting Choices, and Energy Use: an Office Simulation Experiment. *Journal of Environmental Psychology*, *20*(3), 219–237. doi:10.1006/jevp.1999.0169

Veitch, Jennifer A, Stokkermans, M. G. M., & Newsham, G. R. (2011). Linking Lighting Appraisals to Work Behaviors. *Environment and Behavior*. doi:10.1177/0013916511420560

Veitch, Jennifer A. (1997). Revisiting the performance and mood effects of information about lighting and flourescent lamp type. *Journal of Environmental Psychology*, *17*(3), 253–262. doi:10.1006/jevp.1997.0059

Veitch, Jennifer A., Gifford, R., & Hine, D. W. (1991). Demand characteristics and full spectrum lighting effects on performance and mood. *Journal of Environmental Psychology*, *11*(1), 87–95. doi:10.1016/S0272-4944(05)80007-6

Veitch, Jennifer A., & McColl, S. L. (1994). *Full-Spectrum Fluorescent Lighting Effects on People: A Critical Review* (IRC Internal Report No. 659) (pp. 53–111). Ottawa, ON: National Research Council of Canada, Institute for Research in Construction. Retrieved from http://www.nrc-cnrc.gc.ca/obj/irc/doc/pubs/ir/ir659/veitch_mccoll.pdf

Viola, A. U., James, L. M., Schlangen, L. J. M., & Dijk, D.-J. (2008). Blue-enriched white light in the workplace improves self-reported alertness, performance and sleep quality. *Scandinavian Journal of Work, Environment & Health*, *34*(4), 297–306.

Walch, J. M., Rabin, B. S., Day, R., Williams, J. N., Choi, K., & Kang, J. D. (2005). The Effect of Sunlight on Postoperative Analgesic Medication Use: A Prospective Study of Patients Undergoing Spinal Surgery. *Psychosomatic Medicine*, *67*(1), 156–163. doi:10.1097/01.psy.0000149258.42508.70

Wang, N., & Boubekri, M. (2011). Design recommendations based on cognitive, mood and preference assessments in a sunlit workspace. *Lighting Research and Technology*, *43*(1), 55–72. doi:10.1177/1477153510370807

Wang, Na, & Boubekri, M. (2010). Investigation of declared seating preference and measured cognitive performance in a sunlit room. *Journal of Environmental Psychology*, *30*(2), 226–238. doi:10.1016/j.jenvp.2009.12.001

Wargocki, P., & Wyon, D. P. (2012). Providing Better Thermal And Air Quality Conditions In School Classrooms Would Be Cost-Effective. *Building and Environment*, (0). doi:10.1016/j.buildenv.2012.10.007

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. doi:10.1037/0022-3514.54.6.1063

Webb, A. R. (2006). Considerations for lighting in the built environment: Non-visual effects of light. *Energy and Buildings*, *38*(7), 721–727. doi:10.1016/j.enbuild.2006.03.004

Wilkins, A. J., Nimmo-Smith, I., Slater, A. I., & Bedocs, L. (1989). Fluorescent lighting, headaches and eyestrain. *Lighting Research and Technology*, *21*(1), 11–18. doi:10.1177/096032718902100102

Wilson, R., Guilford, J., Christensen, P., & Lewis, D. (1954). A factor-analytic study of creative-thinking abilities. *Psychometrika*, *19*(4), 297–311. doi:10.1007/BF02289230

Wotton, E., & Barkow, B. (1983). An investigation of the effects of windows and lighting in offices. In *Proceedings of the International Daylighting Conference* (pp. 405–411). Washington, DC: American Institute of Architects.

Wu, W., & Ng, E. (2003). A review of the development of daylighting in schools. *Lighting Research and Technology*, *35*(2), 111–124. doi:10.1191/1477153503li072oa

Wunsch, H., Gershengorn, H., Mayer, S. A., & Claassen, J. (2011). The effect of window rooms on critically ill patients with subarachnoid hemorrhage admitted to intensive care. *Critical Care*, *15*(2), 1–10. doi:10.1186/cc10075

Young, H. H., & Berry, G. L. (1979). The Impact of Environment on the Productivity Attitudes of Intellectually Challenged Office Workers. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *21*(4), 399–407. doi:10.1177/001872087902100402

7 APPENDIX

Appendix A: Energy use and cost/benefit calculations

The energy use of the lights in VS319 was monitored during the semester in order to see how much energy the TDDs could save.

It is estimated that with the TDDs, the electric lighting was left off for an additional 5.15 hours a day on average. It should be noted though that, because we were interested in seeing if people would turn the electric lighting on during the day, we checked the room every morning and made sure the lights were off. This would inflate the frequency that the lights were off.

The lights use 2.58kWh of energy.

5.15 hours x 2.58kWh = 13.3kWh/day savings

At an average commercial price of 17c/kWh (Ministry of Economic Development, 2011) savings are $2.26/day.

If the room is used every day of the year except for public holidays then savings would be:

355 days x $2.26 = $802/year

According to Hometech, a typical project like this would cost about $14,500 — though if complications arise the cost could be higher.

A very simple estimate of the payback period puts it at about 18 years ($14,500 ÷ $802).

This is a very basic estimate that ignores net present value, and savings from having to replace the lamps less often, as well as ignoring the inflation of the amount of time the lights were off caused by us turning them off in the morning. More detailed studies have provided similar estimates (Mayhoub & Carter, 2011), suggesting that it is a reasonable ballpark estimate.

Appendix B: Room details

B1: Plans

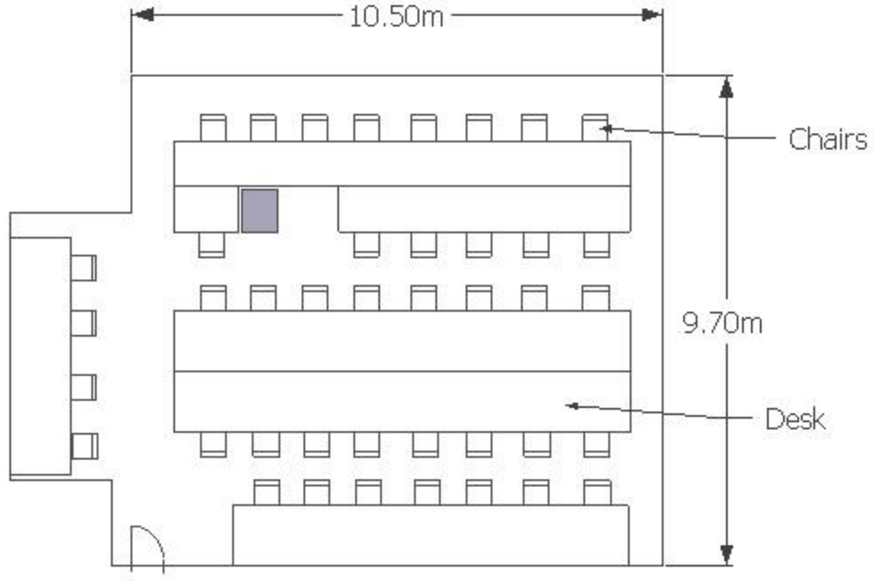


Figure 7-1: VS319 floor plan (1:150)

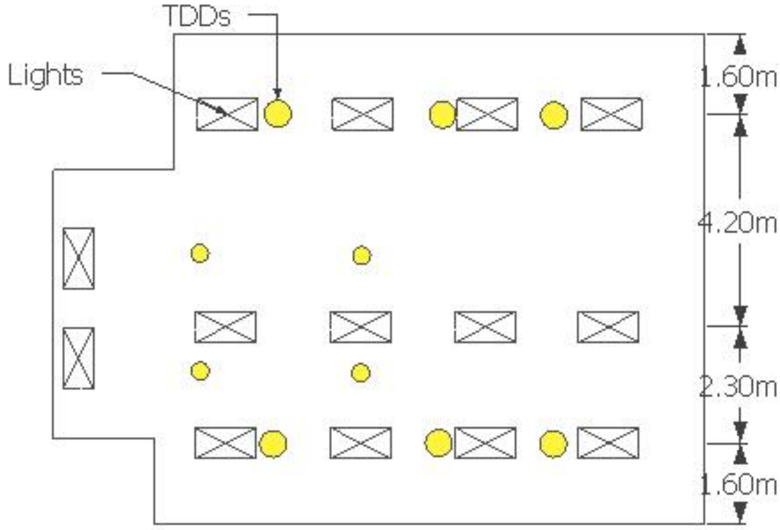


Figure 7-2: VS319 reflected ceiling plan (1:150)

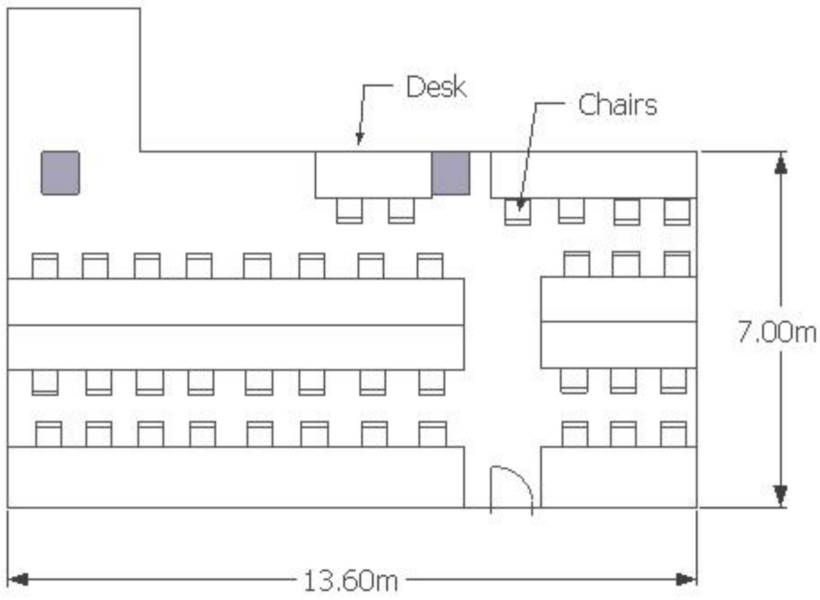


Figure 7-3: VS322 floor plan(1:150)

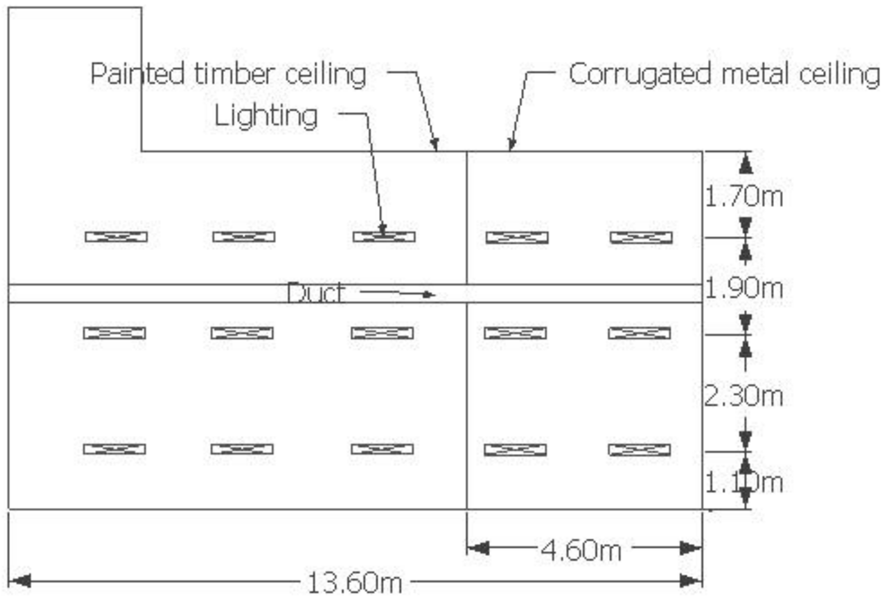


Figure 7-4: VS322 reflected ceiling plan (1:150)

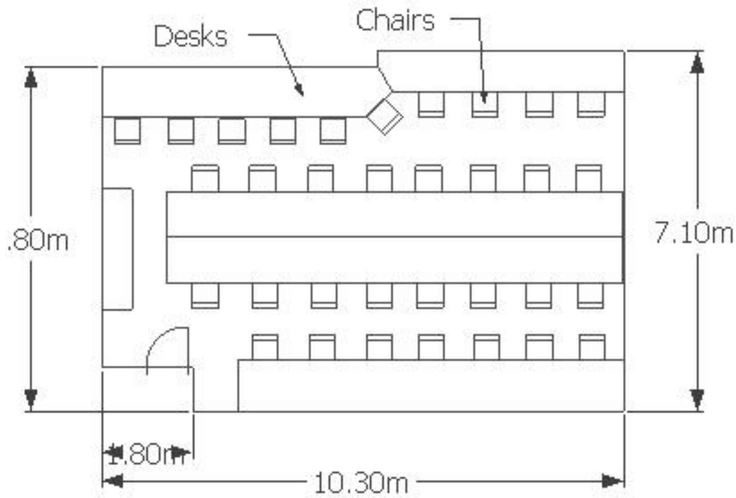


Figure 7-5: VS226 floor plan (1:150)

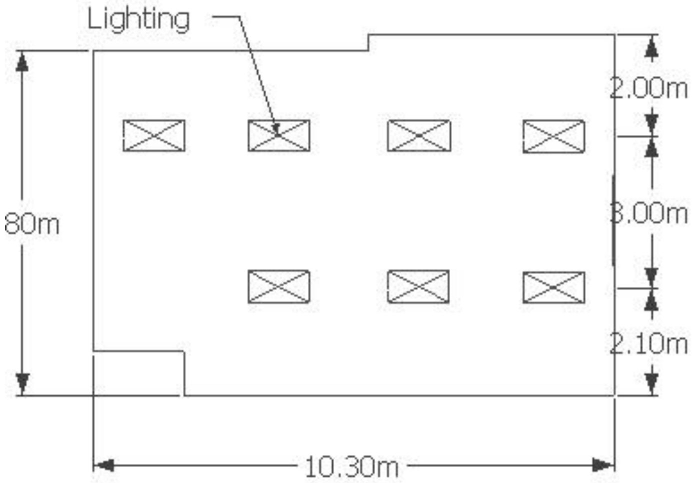


Figure 7-6: VS226 reflected ceiling plan (1:150)

B2: Surface colours/reflectances

Surface colours and reflectances were measured using a ColourMunki spectrometer.

VS319 surfaces	L*	sRGB
Walls	91	(230,229,224)
Back wall	69	(164,168,182)
Floor (carpet)	28	(75,62,63)
Ceiling (corrugated metal)	72	(175,178,178)
Acoustic panels (reflective metal foil)	97	(224,251,252)
Desks	33	(76,77,79)
Chair cushions	36	(45,94,100)
Computers/chair plastic	16	(39,40,42)

Table 7-1: Colours (sRGB) and reflectances (L*) of main surfaces in VS319

VS322 surfaces	L*	sRGB
Walls	91	(230,229,224)
Floor (carpet)	28	(75,62,63)
Ceiling (corrugated metal)	72	(175,178,178)
Ceiling (painted timber)	90	(226,224,218)
Desks	37	(83,88,92)
Chair cushions	36	(45,94,100)
Computers/chair plastic	16	(39,40,42)
Main duct	28	(18,74,108)

Table 7-2: Colours (sRGB) and reflectances (L*) of main surfaces in VS322. Note: surfaces identical to ones in the other rooms are greyed out.

VS226 surfaces	L*	sRGB
Walls	91	(230,229,224)
Floor (carpet)	28	(75,62,63)
Ceiling (painted timber)	90	(226,224,218)
Desks	51	(124,123,113)
Chair cushions	36	(45,94,100)
Computers/chair plastic	16	(39,40,42)
Acoustic panels (reflective metal foil)	97	(224,251,252)

Table 7-3: Colours (sRGB) and reflectances (L*) of main surfaces in VS226. Note: surfaces identical to ones in the other rooms are greyed out.

Appendix C: Environmental measurements

C1: Light levels

Measurements of the illuminances of the electric lighting were taken using a hand-held Minolta T-1 illuminance meter. Measurements were taken at every computer, on both the desktop and computer screen, and were averaged to give the overall light levels for the rooms, as analysis did not show a significant relationship between light level at seat location, and test results (Table 7-7).

Because of the variability of the light provided by the TDDs, LICOR data loggers were employed to monitor the light levels in VS319, taking readings every five minutes. The two sets of light sensors were placed running down the centre of the desks on either side of the room to try and cover as much of the room as possible (Figure 7-7). To estimate what the light levels across the workstations were under TDDs, measurements were taken at the workstations and related to the light levels measured by the sensors, allowing them to be extrapolated to different days. Light levels are given as the average over the tutorials.

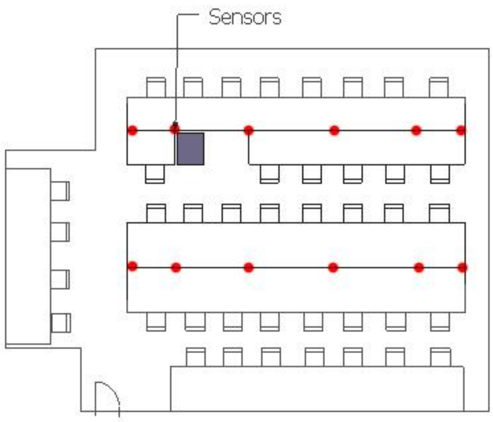


Figure 7-7: Sensor locations

The LICOR sensors were calibrated by Thompson (2012). The hand-held light meter was compared against them to ensure accuracy and consistency (Figure 7-8). Differences were mostly within 30lx, and given that the LICOR readings of the electric light levels fluctuated by up to 20lx, and that an inch or so difference in position of the light meter could change the readings by a similar amount, this was considered to be good agreement.

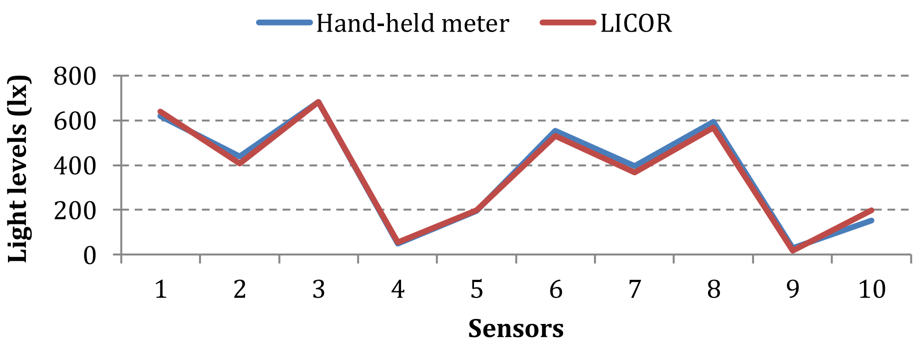


Figure 7-8: Comparison of electric light levels measured by LICOR sensors and the hand-held light meter.

	Horizontal workplane illuminances under TDDs (lx)			
	Performance test 2	Perception test 2	Mood test 3	Perception test 3
Tutorial 1	546	539	594	278
Tutorial 2	585	480	543	166
Tutorial 3	91	384	463	118

Table 7-4: Average illuminances measured on the desktops in VS319 under TDDs during the different tests and tutorial streams.

	Vertical illuminances on screens under TDDs (lx)			
	Performance test 2	Perception test 2	Mood test 3	Perception test 3
Tutorial 1	255	259	286	140
Tutorial 2	282	232	263	82
Tutorial 3	44	193	229	59

Table 7-5: Average illuminances measured on the computer screens in VS319 under TDDs during the different tests and tutorial streams.

	Outdoor illuminances			
	Performance test 2	Perception test 2	Mood test 3	Perception test 3
Tutorial 1	80,000	88,800	88,700	33,000
Tutorial 2	82,750	21,200	87,600	68,000
Tutorial 3	22,500	65,000	78,600	44,000

Table 7-6: Light levels measured outside during the different tests and tutorial streams. Note that they were only measured once during each tutorial to provide a rough idea of what the light levels were. Cloud movement may significantly change the readings.

		Correlations with…	
	n	Horizontal illuminance	Vertical illuminance
Lighting quality	16	-0.04	0.00
Glare problems	16	0.27	0.22
Room attractiveness	16	0.04	-0.06
Illumination	16	-0.34	-0.36
Working memory	36	0.03	0.01
Processing speed	38	-0.08	-0.10
Sleepiness	56	0.08	-0.05

Table 7-7: Correlations between change in response and change in light level at one's seat. Comparison is between overcast and sunny days for perception results, and between 2 days under electric lighting for performance and sleepiness. None are significant.

C2: Air quality

CO_2 levels in the rooms were measured using a hand-held CO_2 meter. The same meter was used for all the measurements to ensure consistency. CO_2 levels varied depending on the location in the room, so multiple readings were taken and averaged to provide overall levels for the rooms.

Outdoor CO_2 levels during the tests were at normal levels for New Zealand, at around 400ppm (BRANZ Ltd, 2007).

	Perception 1	Mood 1	Performance: DSB 1	Mood 2	Performance: Processing 1	Performance 2	Perception 2	Mood 3	Perception 3
VS319: tutorial 1	-	705	665	700		760	850	640	660
VS319: tutorial 2	-	745	665	650	725	610	570	640	610
VS319: tutorial 3	-	780	715	620	700	555	520	590	
VS322: tutorial 1	-	960	1110	1190	1040	1050	1250	1000	1030
VS322: tutorial 2	-	1350	1290	1340	1305	1425	1430	1125	
VS226: tutorial 1	-	805	860	850	810	690	800	700	840
VS226: tutorial 2	-	780	930	940	820	890	800	950	
Outdoors	-	430	400	415	390	385	405	395	410

Table 7-8: CO_2 levels (ppm) measured during the different tests and tutorials

C3: Noise

Noise levels were measured during the performance tests using a hand-held sound meter. Readings were taken every 4 seconds, per standard practice (NZS 6801). Readings were taken from a position roughly in the middle of the room/s.

Before use, the sound meters were calibrated to a 94dB(A) tone, and a correction factor was applied to the readings if necessary.

	Digit Span Backwards test 1	Processing test 1	Performance test 2
VS319	54.4	56.4	61.2
VS322	57.5	60.6	66
VS226	53.25	50.2	60

Table 7-9: Median noise levels measured in the rooms during the performance tests in dB(A)

C4: Air temperature

Air temperature in the rooms was measured using TESTO data loggers, taking readings every five minutes. Temperatures are given as the average temperature measured during the tutorials (Table 7-11, Table 7-12, Table 7-13). The morning and afternoon tutorials in VS226 however were split up due to significant differences between them (>2.5°C).

To ensure consistency and accuracy of the data loggers, they were compared to each other in the same room, and checked against manual measurements from a whirling hygrometer (Table 7-10). They all agreed to within 0.3°C.

	Whirling hygrometer	TESTO 1	TESTO 2	TESTO 3
Air temperature (°C)	22.5	22.6	22.8	22.7

Table 7-10: Comparison of measurement devices for calibration purposes.

	Perception		
	Test 1	Test 2	Test 3
VS319	20.9	19.6	21.3
VS322	21.9	21.7	22.0
VS226-morning	21.3	22.5	22.6
VS226-afternoon	23.7	25.3	23.8

Table 7-11: Mean air temperatures (°C) measured during tutorials with perception tests.

	Performance		
	Digit Span Backwards test 1	Processing test 1	Performance test 2
VS319	21.6	20.4	22.0
VS322	21.8	21.7	22.4
VS226-morning	21.5	22.1	23.3
VS226-afternoon	25.2	23.2	25.7

Table 7-12: Mean air temperatures (°C) measured during tutorials with performance tests.

	Mood		
	Test 1	Test 2	Test 3
VS319	22.3	23.1	22.5
VS322	21.7	22.3	22.1
VS226-morning	22.1	24.0	23.2
VS226-afternoon	24.9	25.6	26.2

Table 7-13: Mean air temperatures (°C) measured during tutorials with mood tests.

Appendix D: Test summaries

Sleepiness

Method: **Karolinska Sleepiness Scale** (Akerstedt & Gillberg, 1990)
Description: Ss place a mark on a 9-point scale rating how sleepy they feel. (1=very alert, 9=very sleepy (fighting sleep)). It is highly correlated to other measures of sleepiness/fatigue/alertness such as EEG (Kaida et al., 2006). **Time:** v. short (it's one simple question)
What it tells us: Subjective sleepiness/fatigue/alertness. Could show if the lighting is helping to stimulate people, or help to maintain alertness in tutorials.
Precedents: Subjective alertness improved under blue enriched white light (17000K) (Mills et al., 2007; Rautkylä et al., 2010; Viola et al., 2008). Alertness higher under blue light than yellow (Lehrl et al., 2007). Daylight is theorised to be able affect people and increase their alertness by stimulating the human circadian system through its higher levels of blue-spectrum light (Figueiro et al., 2011; Webb, 2006; van Bommel, 2006). Links have also been found between sleep quality and illuminance at work (Aries, 2005), and sleep quality and comfort at work (Aries et al., 2010).

Mood

Method: **Positive and Negative Affect Schedule** (Watson et al., 1988)
Description: Questionaire. Ss indicate on scales how much they feel a particular way, e.g. Enthusiastic. 10 items each for both positive and negative affect. Uses 5-point unipolar scales where 1 is 'very slightly or not at all' and 5 is 'extremely'. The average for each set provides a measure of 'Positive mood' and 'Negative mood'. Should be assessed twice, once at the beginning and once at the end, in order to assess the change in mood. **Time:** <5 minutes
What it tells us: How people's mood (positive and negative) changes over time in the room. Can show how well an environment can help maintain a positive mood, or inhibit increase in negative mood. An important note here is that it actually measures Positive and Negative *Activated* affect (feeling alert, and nervous, not relaxed or depressed). It therefore only assesses a section of affect space (Barrett & Russell, 1999). However, this may be appropriate for the task at hand: the scales used for positive affect, for example, are all ones that would seem to be desirable in workers (active, alert, attentive, determined, enthusiastic, excited, inspired, interested, proud, strong). Positive moods have also been linked to a range of psychological effects on people such as improved generosity, more positive perceptions of things, increased pro-social behaviour and improved productivity (Isen, 2001; Russell & Snodgrass, 1991; Veitch et al., 2011).
Precedents: A number of lighting studies have used PANAS: Several have found effects of colour temperature on mood (Knez & Enmarker, 1998; Knez & Kers, 2000; Knez, 1995). Others did not (Boray et al., 1989; Hygge & Knez, 2001). Knez (2001) however has suggested that PANAS might not be a good measure in lighting studies. Other studies have found effects, or at least correlations, of lighting and daylighting on mood (Boubekri et al., 1991; Boyce et al., 2003; Leather et al., 1998; McCloughan et al., 1999; Viola et al., 2008). Sunlight and sunny days also improve positive mood (Cohen, 2011).

Method: **Russell and Mehrabian three-factor semantic differential scale** (Russell & Mehrabian, 1977)

Description:
Questionaire. Ss indicate on paired scales how feel, e.g. Unhappy-happy. There are 3 sets of 6 pairs to assess pleasure, arousal, and dominance (Mehrabian, 1974). Pleasure, arousal, and dominance scores are the average of their sets. Should be assessed twice, once at the beginning and once at the end, in order to assess the change in mood (Boyce et al., 2003).
Time: <5 minutes

What it tells us:
How people's mood changes over time in the room.
Can show how well an environment can help maintain a pleasurable mood, or provide stimulation.
Positive moods have also been linked to a range of psychological effects on people such as improved generosity, more positive perceptions of things, increased pro-social behaviour and improved productivity so it may indirectly suggest many good things (Isen, 2001; Russell & Snodgrass, 1991; Veitch et al., 2011).

Precedents:
A number of lighting studies have used the method to assess mood (Boyce et al., 2006; Newsham et al., 2004; Veitch et al., 1998; Veitch et al., 2008; Veitch et al., 2011; Veitch et al., 1991; Veitch, 1997) and have found effects of lighting on mood.
Several other studies have found effects of colour temperature on mood (Knez & Enmarker, 1998; Knez & Kers, 2000; Knez, 1995). Others did not: (Boray et al., 1989; Hygge & Knez, 2001).
Other studies have found effects, or at least correlations, of lighting and daylighting on mood (Boubekri et al., 1991; Boyce et al., 2003; Leather et al., 1998; McCloughan et al., 1999; Viola et al., 2008). Sunlight and sunny days have also been found to induce positive mood (Cohen, 2011).

Perception of lighting and environment

Method: **Questionnaire**

Description:
Ss are given questionnaires to answer questions about how they feel about the lighting and the space. Questions are generally descriptors on scales, for example, rating how bright/warm/glaring they feel the lighting is. It is also useful to ask how much people like the space, their comfort, and their satisfaction with it.
Time: <5 minutes

What it tells us:
How people are perceiving the lighting and the space, and thus how people see and like the different lightings. Can highlight potential issues such as glare problems.

Precedents:
This is, essentially, the most basic test for dealing with the human response and perception of lighting and the environment. As such, variations are used in many lighting studies (e.g. Aries et al., 2010; Heerwagen & Heerwagen, 1986; Katzev, 1992; Knez, 1995; Miwa & Hanyu, 2006; Moore, Carter, & Slater, 2004; Newsham et al., 2004; Pellegrino, 1999; Slater, Perry, & Carter, 1993)
It is a near certainty that this will show different responses under the different lightings because they are visually different. Indeed, it would be strange, and rather interesting, if people perceived the TDDs and the artificial lighting as the same.

Cognitive performance

There are a many different ways of assessing cognitive performance. Performance effects may only appear on some tasks (e.g. Heschong Mahone Group, 2003a; Knez, 1995; Santamaria & Bennett, 1980) so lone tests may be limited. Multiple tests are often employed in batteries, covering a range of related skills that can be related to typical work (e.g. memory, categorisation, simulated clerical tasks)(e.g. Newsham et al., 2004).

Method: Simple arithmetic/noun underlining test
Description: Ss are given a page of alternating lines of words and successive subtraction problems. They are then given two minutes to go down the page underlining the nouns and solving the math problems. They are scored on the number of correct nouns underlined (24 possible), and the number of correctly solved problems (23 possible). They are not expected to be able to complete the page in 2 minutes (Veitch et al., 1991). **Time:** ~2 minutes
What it tells us: Is a simple measure of basic arithmetic and reading comprehension and speed. These basic skills are felt to be related to the kind of work that is carried out in schools and offices (Boray et al., 1989).
Precedents: No studies that I am aware of have used it when looking at daylighting. Veitch et al. (1991) used it when studying the possible effects of full-spectrum fluorescent lighting. Boray et al. (1989) also used it, finding no effect of colour temperature/spectrum on performance. They suggest that it is unlikely that lamp spectrum will have any effect on simple cognitive performance.

Method: Speed of information processing
Description: Ss are shown a line of 25 random different letters. The letters are light grey on a dark background. They are instructed to read them, from left to right, as fast as possible. They are given 4 seconds to read, and are then asked to write down the last two letters they read (Lehrl et al., 2007). **Time:** ~20 seconds (depends on number of trials, but should be short)
What it tells us: Tells us how fast they are processing information. It has been associated with fluid intelligence and IQ (Lehrl et al., 2007).
Precedents: It has been found that processing speed was faster under blue light than under yellow light (Lehrl et al., 2007).

Method: **Problem solving: Embedded figure task**
Description: Ss are provided with a page of figures. At the top of the page are 5 small 'target' figures identified A-E. Below are 16 larger complex figures in which target figures are hidden. Ss are instructed to identify the target in each of the complex figures (Knez, 1995). Two sheets, for a total of 32 figures are used. It is scored on the number of correct answers. **Time:** 35 minutes
What it tells us: Is considered to be a measure of problem solving ability, and provides a measure of an aspect of cognitive performance (Knez, 1995).
Precedents: Knez (1995, 2001) found that colour temperature of light affected performance. However, Knez & Enmarker (1998) did not find a significant effect. Combined effects of gender and illuminance have also been found (Hygge & Knez, 2001).

Method: **Simple cognitive/clerical performance : Typing task**
Description: Ss have to retype 3 300 word articles from printed copies. The 3 articles are at different font sizes (8pt, 12pt, and 16pt). Ss are scored on speed and accuracy (Newsham et al., 2004). **Time:** 15 minutes
What it tells us: Measures performance on a simple task like that of common office work.
Precedents: The task has been used in a number of lighting studies (Boyce et al., 2006; Newsham et al., 2004; Veitch et al., 2008; Veitch et al., 2011). However, it has been difficult to find any performance effects in experiments (Veitch et al., 2008). Boyce et al. (2006) have suggested that it may not find effects in experimental situations because it is easy for people to work at optimal levels of performance on simple tasks for short periods of time in these kind of situations, unlike in the real world where they may slack off.

Complex cognitive performance

Complex cognitive performance is assessed using three different methods described below, performed sequentially (Newsham et al., 2004). The method has been used in a number of lighting studies (Boyce et al., 2006; Newsham et al., 2004; Veitch et al., 2008; Veitch et al., 2011). However, it has had difficulty finding any performance effects in experiments (Veitch et al., 2008), and Boyce et al. (2006) have suggested that it may not find effects in experimental situations because it is easy for people to work at optimal levels of performance on simple tasks for short periods of time in these kind of situations, unlike in the real world where they may slack off.

Method: **Article categorisation**
Description: Ss are given a 40-60 word summary of an article, as well as 4 possible categories that it could be placed in (Newsham et al., 2004). They then have to select the correct category/s. Time taken and correctness of categorisation are measured. **Time:** Variable – but together the tasks should take about 40 minutes total.
What it tells us: Provides a measure of ability to perform more complex cognitive tasks that are alike those encountered in everyday life.

Method: Summary evaluation

Description:
Ss are given the full 300 word article to read, along with the summary. They are asked to rate how accurately the summary represents the article, the correctness of the grammar, and how good the writing is (Newsham et al., 2004).

Time: Variable – but together the tasks should take about 40 minutes total.

What it tells us:
Provides a measure of effects on cognitive judgements.

Method: Summary extraction

Description:
Ss are given the full 300 word article to read. They are then required to select the 4 most important sentences in the article to summarise it (Newsham et al., 2004). Their speed is measured.

Time: Variable – but together the tasks should take about 40 minutes total.

What it tells us:
Provides a measure of ability to perform more complex cognitive tasks that are alike those encountered in everyday life.

Short-term memory

Method: Article reading

Description:
Ss are given two minutes each to read two short articles (300-350 words). They then answer simple questions about them in a limited time (10 questions for each article, 2 minutes each) without referring back to the articles (Na Wang & Boubekri, 2010). Score is the average number of correct answers.

Time: ~8 minutes

What it tells us:
It provides a measure of short-term memory in a situation that is alike that of normal office work (Na Wang & Boubekri, 2010).

Precedents:
Wang & Boubekri (2010) found some significant effects based on where people were sitting in the room, but could not link it to daylighting.
Other studies looking at short term memory using other methods have found some effects of colour temperature (Knez, 2001), and no effect of colour temperature (Knez & Enmarker, 1998; Knez, 1995). Effects have also been found from light level (Hygge & Knez, 2001).
Another study found a link between level of daylight and performance on a different working memory task (Heschong Mahone Group, 2003b).

Method: Free recall of words

Description:
Ss are presented with a series of 16 words. The words are displayed separately, for 1.5 seconds each. After the list is complete, the subjects are asked to write down all the words they can remember from the list (Knez, 1995). Three different lists with positive, neutral, and negative hedonic tone may be used. This is because different moods may affect recall of different words differently – e.g. people in a positive mood may remember positive things more easily (Russell & Snodgrass, 1991).

Time: <5 minutes

What it tells us: It provides a slightly more abstract measure of short term memory that can be linked with effects of mood. Performance here could indicate effects on memory or effects on mood. A potential issue is that the effects may be limited to only words with the appropriate hedonic tone – e.g. TDDs may improve positive mood and thus improve recall of positively toned words, but not neutral or negative words. This would significantly limit the potential value of the effect.
Precedents: Studies are inconsistent, and have found both some effects of colour temperature (Knez, 2001), and no effect of colour temperature (Knez & Enmarker, 1998; Knez, 1995). Effects have also been found from light level (Hygge & Knez, 2001). Another study found a link between level of daylight and performance on a working memory task (Heschong Mahone Group, 2003b) Effects based on mood may have difficulty finding effects if they are only measured once, due to the confounding effects of external factors such as how they felt coming into tutorial (Boyce et al., 2003).

Method: **Working memory: Digit Span Backwards**
Description: Traditionally it is done verbally, however a visual computer-based variant such as that used by the Heschong Mahone Group (2003a) would be more appropriate for this study. Ss are given a string of numbers, one after another, each number presented 1 second apart. They are then asked to repeat the numbers backwards. They are first given 2 strings of 3 numbers, and then a set of 4 numbers, and so forth progressing up to strings of 9 numbers. They are scored on the longest set of strings that they can accurately remember (Heschong Mahone Group, 2003b). **Time:** A few minutes (varies depending on the number of strings completed)
What it tells us: It is a widely recognised measure of working memory and attention – both important aspects of performance (Heschong Mahone Group, 2003b).
Precedents: Using this method, a study found a link between level of daylight and performance on a short term memory task (Heschong Mahone Group, 2003b) Other studies looking at light and short term memory are inconsistent, and have found both some effects of colour temperature (Knez, 2001), and no effect of colour temperature (Knez & Enmarker, 1998; Knez, 1995). Effects have also been found from light level (Hygge & Knez, 2001).

Method: **Long-term recall and recognition of a text**
Description: Ss read a 7 page text about a subject (e.g. an ancient culture)(Knez, 1995). Later on in the study, after about 90 minutes, they are asked to answer 6 general questions (testing recall) and 18 multiple choice questions (testing recognition) about the text (Knez, 1995). **Time:** 35 minutes (reading), 90 minute gap, 20 minutes (questions)
What it tells us: It provides a measure of longer term memory with a task that is similar to ones carried out, for example, in educational facilities.
Precedents: Several studies have found that colour temperature effects performance (Knez & Kers, 2000; Knez, 1995, 2001). Another, however, found no significant effect (Knez & Enmarker, 1998). No daylighting precedents.

Creative performance

Method: **Guilford's alternate uses test**
Description: Ss are asked to come up with as many different uses for, say, a brick in a certain time frame. They can be assessed on the number of uses they came up with, and the variety of different uses (Goncalo et al., 2010). It was originally created in 1954 as the Unusual uses or Brick uses test (Wilson et al., 1954). **Time:** 10 minutes
What it tells us: Assesses divergent thinking – the ability to think of different solutions to a problem. Is a measure of creativity (Goncalo et al., 2010). May be difficult to have subjects repeat.
Precedents: No studies that I am aware of assess the effects of lighting on creativity. However, positive mood (which can be enhanced through lighting) has been linked to improved creative thinking (Isen et al., 1987).

Method: **Ward's measure of structured imagination**
Description: Ss are asked to draw a creature from an alien planet that is very different from earth. They are assessed on how much their creature differs from earth norms. Scoring is done independently by two raters whose scores can then be compared to make sure that they mostly agree (Goncalo et al., 2010). **Time:** 7 minutes
What it tells us: Measures ability to come up with novel ideas (Goncalo et al., 2010). This means that it may have problems with being administered to subjects a second time.
Precedents: No studies that I am aware of assess the effects of lighting on creativity. However, positive mood (which can be enhanced through lighting) has been linked to improved creative thinking (Isen et al., 1987).

Method: **Torrence tests of creative thinking**
Description: The Torrence tests are a series of exercises, like those of the previous methods, designed to assess a range of aspects of creative thinking (Dow, 2003). It is the most commonly used measure of creativity, however it is noted for difficult scoring that generally requires one to send the results to professionals to be scored (Clapham, 2004). This makes it less accessible and easy to use. **Time:** Unknown, but the number of tests means it is likely much longer than the other creativity tests.
What it tells us: Measures creativity.
Precedents: No studies that I am aware of assess the effects of lighting on creativity. However, positive mood (which can be enhanced through lighting) has been linked to improved creative thinking (Isen et al., 1987).

Motivation

Method: **NRC Conveyer belt task**
Description: Symbols travel along a computer screen, passing through a box. Ss press a button as fast as possible to remove certain target symbols when they enter the box. The speed of the symbols gradually increases and Ss are instructed to stop when they can no longer handle it (Newsham et al., 2004). Persistence is measured by the maximum speed the Ss reach before they give up (Boyce et al., 2006a). **Time:** 10 minutes
What it tells us: Provides a measure of motivation, defined as the willingness to persist at a difficult task – something which is useful when people are carrying out difficult tasks. Note: may be confounded with skill at task.
Precedents: The task has been used in a number of lighting studies (Boyce et al., 2006; Newsham et al., 2004; Veitch et al., 2008; Veitch et al., 2011). It has found effects from lighting control (Boyce et al., 2006a), but not from lighting design (Boyce et al., 2006a). It has also been found that people in a positive mood tend to be more motivated (Veitch et al., 2011).

Visual performance

Method: **Landolt ring test**
Description: Ss are given a page with a 10x10 grid of rings with gaps in them in one of the eight cardinal directions (Boyce, 1974). They must mark all of the rings with a gap in a specific direction, as well as the rings with no gaps in them, with a red pen. Both speed and accuracy are assessed (Boyce, 2003). The size of the gap and the contrast of the rings can be adjusted to change the task difficulty. **Time:** ~5 minutes for 3 trials (Boyce, 1974)
What it tells us: Provides a measure of performance on a visual task that is less dependent on cognitive skills such as reading ability.
Precedents: The Landolt ring test has been used extensively in lighting research (Boyce, 2003). Visual performance is largely dependent on the light levels and the contrast, and it tends to plateau once a certain minimal level is reached (Boyce et al., 2003). The spectral properties of daylight are more valuable for tasks requiring fine colour discrimination, and near-threshold (i.e. very difficult to see) tasks. For these reasons, daylight is not inherently better than artificial light for most visual tasks (Boyce et al., 2003). Thus, it is unlikely that this or other simple visual performance tests will find any significant effect from TDDs for task difficulties that are applicable to everyday work.

Colour discrimination

Method: Farnsworth-Munsell 100-hue test
Description: Ss are given 85 coloured 'caps' in 4 rows. They are then required to arrange the caps in order of hue. They are assessed on the number of errors they make (Boyce & Cuttle, 1990). **Time:** ~15 minutes (Hawes et al., 2012). However various issues make it take much longer. Specifically, the fact that it requires specialised equipment, so only 1 subject can take it at once. Multiple sets of the equipment would be prohibitively expensive, as a single set costs hundreds of dollars. Furthermore, recording and analysing the results can also take a substantial amount of time (Hidajat et al., 2004).
What it tells us: Ability to accurately discriminate colours under different lights. Could show if TDDs are good for tasks that require fine colour discrimination.
Precedents: The test is widely used to test colour discrimination (Kinnear & Sahraie, 2002), and has been used in lighting research before (Boyce & Cuttle, 1990). Daylight is known to be good for fine colour discrimination (Boyce et al., 2003), moreover, the 100-hue test is designed to be run under light like that of daylight (Boyce & Cuttle, 1990). It is thus reasonably safe to predict that the light from the TDDs will be good for performance on the test. Value of any effects here would be limited to situations that require good colour discrimination – like carpet weaving (Boyce et al., 2003).

Seating preferences

Method: Reported preference
Description: Ss are given a picture of the room, and asked to mark where they would prefer to sit. Combining the responses can map out where the preferred seating locations are (Wang & Boubekri, 2010). **Time:** Short, and less important. As the method does not require Ss to be *in* the room, it can be run separately from the tests in the tutorials.
What it tells us: Where people think they would prefer to sit if they had the option. May show what people like in their environment, which could be affected by the TDDs.
Precedents: Wang & Boubekri (2010) used it when studying seating location and performance. They found that the preferred locations were not the same as the ones with the best performance, and could not directly link either to daylighting. Other studies have found that people generally prefer to sit by windows (Boyce et al., 2003; Kilic & Hasirci, 2011; Shemirani et al., 2011). Of course, TDDs are not the same as windows.

Method: **Observing behaviour**
Description: Observe the room and see where people first go to sit. **Time:** Requires prolonged observation. However, it can be done separately from the other tests, as it does not require interaction with Ss.
What it tells us: The preferred seats. This may be able to be correlated with local environmental conditions, showing what people like in their environment.
Precedents: Observational studies have found that people generally prefer to sit by windows (Kilic & Hasirci, 2011; Shemirani et al., 2011). Wang & Boubekri (2010) however found that the preferred locations were not the same as the ones with the best performance, and could not directly link either to daylighting. Could be difficult to find anything as people's behaviour is likely to be heavily influenced by other factors such as where other people are sitting.

Health

Method: **SF-36 Questionaire**
Description: The SF-36, or the shorter SF-12 questionnaire are questionnaires that ask the subjects questions about their health. Assesses both physical and mental health. (Mills et al., 2007). **Time:** Depends on the length of questionaire. However to assess the effects of the environment on health the Ss must be exposed to it for a prolonged amount of time – such as in an office building where they work 9-5. However, the students will only be in the computer room intermittently, so effects on health cannot be assessed.
What it tells us: How healthy the Ss are, which could then be associated with their environment. Healthier workers are both good for society, as they cost less in healthcare, and more productive as they would be absent less.
Precedents: Using SF-36, Mills et al. (2007) found that blue-enriched white light in offices could provide improvements in vitality and mental health. Other studies have also found effects of daylighting on health: Aries et al. (2010) found that reducing discomfort at work, such as by providing good daylighting and views, could also improve sleep quality. Another study found a positive correlation between sleep quality and the level of vertical illuminances at work (Aries, 2005). In Germany it was found that workers closer to windows had less health problems (Çakir & Çakir, 1998). Other studies have found that workers that get less daylight get more headaches (Boyce et al., 2003; Wotton & Barkow, 1983).

Stress

Stress has linked to many undesirable effects in people, such as reduced job performance, poor health, higher turnover, and increased absenteeism (Boyce et al., 2003).

Investigating subjective stress is much the same as investigating health as discussed above. Measures of stress are generally looking at job stress, and are used in field studies (e.g. Fostervold & Nersveen, 2008) that require workers to say how much of a problem they have with various symptom of stress. Thus, like health surveys, they require long term exposure to the environmental conditions being studied. Thus, stress may not be assessed in this study.

Absenteeism

Method: **Record Absenteeism**
Description: One could get records of the amount of absenteeism in tutorials in the room and compare them with and without the TDDs. This would, obviously, require the different lighting conditions to be active over prolonged periods of time. For example, having artificial lighting on for the first half of the semester, and the changing to TDDs for the second half. **Time:** NA – just need to collect records
What it tells us: If the lighting can reduce absenteeism, which would be potentially very valuable, and could also indicate health effects.
Precedents: Absenteeism is an oft used indirect method for examining productivity (Heschong Mahone Group, 2003b). It has been used in lighting research before, however direct links between lighting and absenteeism have not been clearly demonstrated: In their study of classrooms, the (Heschong Mahone Group, 2003a) did not find any correlation between the environmental characteristics of the classrooms and absenteeism, and suggested it may be because students are often absent for a range of reasons other than health. Another study found that workers in buildings with high amounts of glazing had more eyestrain and absenteeism (Wotton & Barkow (1983), in Galasiu & Veitch (2006)). Some case studies also suggest a possible effect of environment, with two examples of organisations which moved into new daylit office buildings and reported 15% less absenteeism (Romm & Browning, 1994). Difficulties come, however, in the fact that effects (if they exist) are likely to be covered up by the other reasons that students have for being absent. Anecdotally, the primary factors in absenteeism are a) students just not bothering to come to tutorial – often because they haven't done any work, b) that absenteeism increases the later it is in the semester, and c) that absenteeism will dramatically increase when there is an assignment due soon.

Visual comfort

Method: **Eye discomfort scale**
Description: Ss are asked to indicate the intensity of various symptoms of eye discomfort on a scale of 0-4. Symptoms are: smarting, itching, gritty feeling, aches, sensitivity to light, redness, teariness, and dryness (Newsham et al., 2004). **Time:** Very short – maybe a minute or two
What it tells us: Subjective feelings of eye discomfort. Indicates whether or not they are suffering from eyestrain.
Precedents: The test has been used in a number of lighting studies (Boyce et al., 2006a; Newsham et al., 2004). A field study found that workers in buildings with 68% glazing suffered more eyestrain than those in ones with 11% glazing (Galasiu & Veitch, 2006). It has also been found that blue enriched white light can improve eye discomfort in office workers (Viola et al., 2008). It should be noted however that effects of eye discomfort are only likely to be seen in the long term – it is often used in field studies and Ss are normally asked how often they experience the symptoms (Newsham et al., 2004). While frequency is non-applicable to a short study, it should be noted that the studies of Newsham et al. (2004) covered a period of a day, and found few symptoms. It is thus unlikely that the test would find anything in such a short study as this.

Attitude towards university/work etc.

Method: **Questionnaire**
Description: Survey Ss on how they feel about university, the building, the facilities and so forth. Possibly ask how sustainable they feel the university is, as the sustainability argument is a common reason for promoting daylighting. **Time:** Unknown, should be short given a small number of questions – maybe a minute or two?
What it tells us: Whether or not TDDs can make people feel better about not just the space, but about other things like the building in general and the organisation they're working with. If this was the case, then TDDs could be said to improve people's image of organisations and their buildings - which could make them economically useful.
Precedents: Effect may be generated *if* TDDs improved positive mood. Studies have found that improving mood by, for example, providing a pleasant environment can also enhance how much people like other, unrelated, things (Russell & Snodgrass, 1991). It may be difficult to study though as it could be heavily confounded with other factors, such as how the Ss are doing in their courses. It may be problematic to survey repeatedly, as people could get set in their opinions by answering the questions, or will discuss the survey with each other, which could then influence their answers.

Ethics

Method: **Defining Issues Test** (Rest & Narvaez, 1998)
Description: Ss are given a scenario with a moral dilemma. They are given a series of issues that may be relevant to how they think through the dilemma, and are asked to rate how important they feel each issue is when thinking about the problem. For example (Rest & Narvaez, 1998), a dilemma may be whether or not a reporter should report a story about some minor misdemeanour an election candidate committed when they were a teenager. One of the issues may be whether or not the public has a right to know everything about political candidates. The Ss would have to rate how important they feel that issue is to the problem. They are then asked to decide what the correct course of action is. There are 5 dilemmas in the test. **Time:** Unknown, but likely long – estimate over 30 minutes at least
What it tells us: What people think about certain moral issues and what schemas they use to decide (Narvaez & Bock, 2002). It may be able to show if people's moral judgement can be changed by lighting, which would be interesting, and may suggest that it could positively influence people's behaviour.
Precedents: To the author's knowledge, no studies have looked at the effects of lighting on moral judgement. The prompt for this came from a study that found that people who were asked to remember unethical deeds perceived a room as being darker, and that they expressed greater preference for light (Banerjee et al., 2012). Studies of positive affect have found that it can affect potentially related things like generosity, 'liking' of things, and acceptability of political messages (Russell & Snodgrass, 1991). This raises the possibility that lighting could affect moral judgments as well.

Conflict resolution

Method: **Hypothetical conflict resolution**
Description: Ss read a scenario describing a workplace conflict. They are then given 5 different possible strategies of conflict resolution, and are asked to rate how likely it is that they would use that strategy. They are also asked to rank the strategies (Boyce et al., 2006a). The five strategies are: compromise, accommodation, competition, avoidance, and collaboration (Baron et al., 1992). Multiple scenarios may be given (Newsham et al., 2004). **Time:** 30 minutes (using 5 scenarios) – presumably shorter (~5min) if only using 1 scenario like Baron et al. (1992).
What it tells us: Whether or not TDDs can affect people's conflict resolution behaviour in a way that may be more productive for organisations.
Precedents: This method has been used in several lighting studies (Baron et al., 1992; Boyce et al., 2006a; Newsham et al., 2004). Baron et al. (1992) found that Ss had a greater preference for resolving conflicts with collaboration under warm white light than under cool white light. Studies of positive affect have also found that improved positive mood in people can also affect people's responses (Baron et al., 1992). TDDs may affect it if they affect positive mood.

Appendix E: Tests and survey forms

E1: Lighting quality

The questions (except for #7) were from the Light Quality survey used by the National Research Council Canada (Veitch & Newsham, 2000). Details such as the labels on the scales (e.g. "very satisfied", "fairly satisfied" etc.) were provided to us by Dr. Veitch and Dr. Newsham.

Question #7 was added by the researchers for the purposes of this study.

Lighting Quality

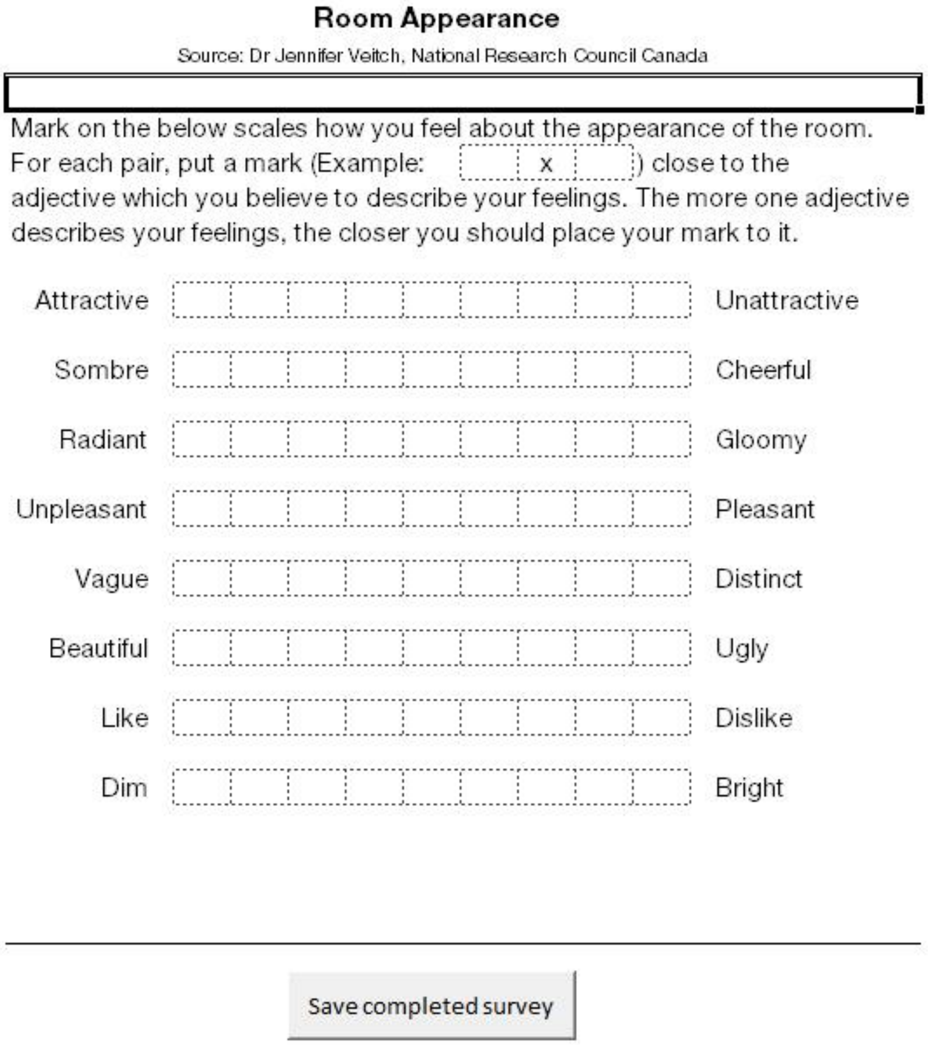
Questions from: Veitch, J. A., & Newsham, G. R. (2000). Exercised Control, Lighting Choices, and Energy Use: an Office Simulation Experiment. *Journal of Environmental Psychology*, 20(3), 219–237.

Considering the lighting in this room, please place a mark (Example: [x]) under the descriptor that best describes your judgement for each question.

1. Overall, how satisfied are you with the lighting at your work space?	not at all satisfied	not very satisfied	neither satisfied nor dissatisfied	fairly satisfied	very satisfied
2. Rate the lighting available to you for reading.	poor	not very good	neutral	pretty good	excellent
3. Rate your workstation on the amount of light for the work you do.	poor	not very good	neutral	pretty good	excellent
4. Lighting at my desk hindered me from doing my job well.	strongly disagree	disagree	neutral	agree	strongly agree
5. How much do the reflections in the computer screen bother you?	not at all	a little	moderately	considerably	extremely
6. How much does the glare bother you?	not at all	a little	moderately	considerably	extremely
7. Rate the daylight available to you	poor	not very good	neutral	pretty good	excellent
8. How do you rate the acceptability of the lighting in this office?	completely unacceptable		barely acceptable		completely acceptable

Figure 7-9: Lighting quality survey

E2: Room appearance



Room Appearance

Source: Dr Jennifer Veitch, National Research Council Canada

Mark on the below scales how you feel about the appearance of the room. For each pair, put a mark (Example: [| x |]) close to the adjective which you believe to describe your feelings. The more one adjective describes your feelings, the closer you should place your mark to it.

Attractive										Unattractive
Sombre										Cheerful
Radiant										Gloomy
Unpleasant										Pleasant
Vague										Distinct
Beautiful										Ugly
Like										Dislike
Dim										Bright

Save completed survey

Figure 7-10: Room appearance survey

E3: Mood

Russell and Mehrabian Semantic Differential Scale

Source: Mehrabian, A. (1974). An approach to environmental psychology. Cambridge: M.I.T. Press.

Each pair of words below describes a feeling dimension. Some of the pairs might seem unusual, but you may generally feel more one way than the other. So, for each pair, put a mark (Example: [ ] [x] [ ]) close to the adjective which you believe to describe your feelings better *at this moment*. The more appropriate that adjective seems, the closer you put your check mark to it.

Unhappy										Happy
Sleepy										Wide-awake
Stimulated										Relaxed
Dull										Jittery
Relaxed										Bored
Contented										Melancholic
Calm										Excited
Pleased										Annoyed
Despairing										Hopeful
Frenzied										Sluggish
Unsatisfied										Satisfied
Aroused										Unaroused

Save completed survey

Figure 7-11: Mood survey. From Mehrabian (1974).

E4: Pre-test control questions

General Questions

Please answer the below questions by placing a mark in the correct box Example: [x]

1. Have you drunk coffee (or another caffeinated beverage) today? Yes: [ ] No: [ ]

2. Have you had lunch? Yes: [ ] No: [ ]

3. Describe how you feel at this moment on the below scale:

Very alert		Alert		Neither alert nor sleepy		Sleepy (but not fighting sleep)		Very sleepy (fighting sleep)

Figure 7-12: Control questions given at the start of each test. Note: "lunch" is replaced with "breakfast" in the morning tutorials.

E5: Demographic survey

Demographic Information

Please answer the below questions by placing a mark in the correct box Example: [x]

1. What is your age group? Below 30: [ ] Above 30: [ ]

2. What is your gender? Male: [ ] Female: [ ]

3. Is English your native language? Yes: [ ] No: [ ]

4. Do you have any visual impairments? Mark the appropriate options:

Require corrective lenses: [ ]

Colour blind: [ ]

Figure 7-13: Demographic questions

E6: Lighting beliefs survey

Lighting Effects Questionnaire

Source: Veitch, J. A., & Gifford, R. (1996). Assessing beliefs about lighting effects on health, performance, mood and social behaviour. Environment and Behavior, 28(4), 446–470.

Say how much you agree or disagree with the following statements about lighting by placing a mark on the below scales.
If you feel that you cannot answer the question, then you can mark "Don't know"

1. The quality of light wherever I am is important to my well-being	Disagree	Agree	Don't know
2. You cannot get skin cancer from working under fluorescent lights	Agree	Disagree	Don't know
3. Sunny days make me happy	Disagree	Agree	Don't know
4. Bright lights are stimulating; they make me feel energetic	Disagree	Agree	Don't know
5. I get eyestrain from working under fluorescent lights	Disagree	Agree	Don't know
6. Incandescent lights are relaxing	Disagree	Agree	Don't know
7. I learn equally well in a room with any kind of lights	Agree	Disagree	Don't know
8. Bright light at work does not improve my morale	Agree	Disagree	Don't know
9. The brighter the light, the more work I accomplish	Disagree	Agree	Don't know
10. Glittery, dazzling lights rarely make me dizzy	Agree	Disagree	Don't know
11. Bright, harsh fluorescent lighting can make me feel tense	Disagree	Agree	Don't know
12. The quality of light in my workplace is irrelevant to my job satisfaction	Agree	Disagree	Don't know
13. Natural daylight indoors improves my mood	Disagree	Agree	Don't know
14. Bright lights in grocers and pharmacies don't make me buy more	Agree	Disagree	Don't know
15. It makes no difference to me what kind of lighting is in a room	Agree	Disagree	Don't know
16. Soft, diffuse light is soothing	Disagree	Agree	Don't know
17. Pregnant women should avoid exposure to fluorescent lighting	Disagree	Agree	Don't know
18. My vision never becomes blurred when the lights are very bright	Agree	Disagree	Don't know
19. Glaring lights give me headaches	Disagree	Agree	Don't know
20. I rarely use warm-coloured lighting to help me relax	Agree	Disagree	Don't know
21. Reading under dim light doesn't damage your vision	Agree	Disagree	Don't know
22. Fluorescent lights are bad for your health	Disagree	Agree	Don't know
23. If I want to create an intimate setting, I dim the lights	Disagree	Agree	Don't know
24. Bright light makes people talk louder	Disagree	Agree	Don't know
25. Fluorescent light seldom gives me a headache	Agree	Disagree	Don't know
26. If a restaurant is very brightly lit, I will leave soon after I've finished eating	Disagree	Agree	Don't know
27. Lack of sunlight in winter does not bother me	Agree	Disagree	Don't know
28. Incandescent lighting in a room helps me to pay attention to the speaker	Disagree	Agree	Don't know
29. I do my best work in places that are lit using natural daylight	Disagree	Agree	Don't know
30. Humming noise from fluorescent lights usually does not distract me	Agree	Disagree	Don't know
31. Bright lights rarely make me feel excited and full on anticipation	Agree	Disagree	Don't know

Figure 7-14: Lighting Beliefs survey (from Veitch & Gifford, 1996). Note that one question about full-spectrum fluorescent lighting was removed as it was not relevant to this study and we did not want to have to explain to the students what full-spectrum lighting was.

Appendix F: Analysis/Results

F1: Perception results

			Lighting					
			Lighting quality		Glare		Daylight Quality	
	Room	*n*	mean	SD	mean	SD	mean	SD
Test 1	VS226	17	2.19	0.89	1.59	1.28	0.47	0.87
	VS319	49	2.28	0.76	1.42	1.24	0.27	0.64
Test 2	VS322	21	2.65	0.58	1.43	1.10	0.48	0.93
	VS226	14	2.36	0.64	1.36	0.91	0.21	0.58
	VS319	23	2.90	0.59	1.57	1.25	1.61	1.44
Test 3	VS322	4	2.35	0.55	1.38	0.48	0.00	0.00
	VS226	11	2.22	0.58	0.64	0.71	0.09	0.30
	VS319	22	2.62	0.78	1.07	0.98	1.86	1.32

Table 7-14: Lighting survey results

			Room Appearance			
			Attractiveness		Illumination	
	Room	*n*	mean	SD	mean	SD
Test 1	VS226	17	3.38	1.78	4.43	1.46
	VS319	49	3.85	1.18	4.34	1.33
Test 2	VS322	21	3.34	1.36	4.17	1.25
	VS226	14	3.46	1.72	4.38	1.15
	VS319	23	4.91	1.31	5.75	1.07
Test 3	VS322	4	3.65	1.91	4.50	0.84
	VS226	11	4.02	1.60	4.64	1.31
	VS319	22	5.13	1.29	5.24	1.16

Table 7-15: Room appearance survey results

F2: Performance results

	Room	*n*	Digit Span Backwards Average	SD
Test 1	VS322	20	5.28	0.90
	VS226	21	5.26	1.62
	VS319	45	5.10	1.18
Test 2	VS322	21	4.60	1.45
	VS226	15	5.08	1.94
	VS319	24	4.54	1.22

Table 7-16: Digit Span Backwards test results

	Room	*n*	Processing Speed Average	SD
Test 1	VS322	19	14.87	3.34
	VS226	21	14.67	3.58
	VS319	45	15.32	3.00
Test 2	VS322	16	14.86	2.83
	VS226	12	14.89	3.49
	VS319	23	14.36	2.62

Table 7-17: Processing speed test results

Summary output of regression model for Digit Span Backwards

Regression Statistics	
Multiple R	0.20
R Square	0.04
Adjusted R Square	0.01
Standard Error	1.34
Observations	143

ANOVA

	df	SS	MS	F	Significance F
Regression	4	10.81	2.70	1.51	0.20
Residual	138	247.63	1.79		
Total	142	258.44			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	3.10	0.20	15.52	0.00	2.71	3.49	2.71	3.49
room 226	0.27	0.33	0.81	0.42	-0.38	0.91	-0.38	0.91
room 322	0.07	0.33	0.20	0.84	-0.59	0.72	-0.59	0.72
2nd round	-0.47	0.32	-1.47	0.14	-1.09	0.16	-1.09	0.16
TDDs	-0.09	0.46	-0.20	0.84	-1.01	0.82	-1.01	0.82

F3: Mood results

			Δ Pleasure		Δ Arousal	
	Room	***n***	**Average**	**SD**	**Average**	**SD**
Test 1	**VS322**	16	0.15	1.13	0.28	0.76
	VS226	12	-0.24	0.80	0.04	0.76
	VS319	34	-0.52	1.17	-0.06	0.73
Test 2	**VS322**	4	0.29	0.92	0.46	0.37
	VS226	18	-0.08	1.61	0.19	0.65
	VS319	23	0.19	1.49	0.14	1.25
Test 3	**VS322**	10	0.05	1.28	0.35	0.78
	VS226	11	-0.20	0.73	0.02	0.54
	VS319	14	0.38	1.33	0.11	0.66

Table 7-18: Results for change in mood

F4: Sleepiness results

	Room	*n*	Mean	SD
Test 1a	**VS319**	49	5.10	1.69
	VS226	17	5.12	1.50
Test 1b	**VS319**	44	4.84	1.88
	VS322	24	5.25	1.73
	VS226	20	4.95	1.90
Test 1c	**VS319**	45	5.41	2.03
	VS322	25	5.16	1.97
	VS226	23	5.17	1.85
Test 1d	**VS319**	45	4.96	1.85
	VS322	20	4.85	1.66
	VS226	21	5.95	1.63
Test 1e	**VS319**	42	5.00	1.99
	VS322	16	5.19	2.07
	VS226	22	4.60	1.64
Test 2a	**VS319**	24	4.96	1.57
	VS322	21	5.10	1.84
	VS226	15	5.53	1.96
Test 2b	**VS319**	23	5.57	1.88
	VS322	21	4.71	1.31
	VS226	14	4.07	1.82
Test 2c	**VS319**	22	4.45	1.79
	VS322	13	4.38	2.29
	VS226	21	4.70	1.95
Test 2d	**VS319**	22	4.55	1.47
	VS322	4	4.00	1.15
	VS226	11	4.73	1.79

Table 7-19: Sleepiness results. Tests 1(a-e) are in the first half of the semester, and Tests 2(a-d) are in the second half.

F5: Psychological factors

	Lighting Quality	Glare Problems	Daylight Quality	Room Attractiveness	Illumination
The quality of light wherever I am is important to my well-being	0.10	0.26	0.11	-0.20	-0.20
You cannot get skin cancer from working under fluorescent lights	0.35	-0.02	0.43	0.42	0.33
Sunny days make me happy	0.12	-0.17	0.15	-0.04	0.38
Bright lights are stimulating; they make me feel energetic	-0.08	-0.22	-0.16	-0.06	0.22
I get eyestrain from working under fluorescent lights	0.43	-0.27	0.31	0.50	0.37
Incandescent lights are relaxing	0.22	-0.29	0.08	0.12	0.11
I learn equally well in a room with any kind of lights	-0.10	-0.04	-0.08	0.04	0.19
Bright light at work does not improve my morale	-0.18	0.28	-0.14	0.05	0.10
The brighter the light, the more work I accomplish	-0.01	-0.02	0.17	0.02	0.19
Glittery, dazzling lights rarely make me dizzy	-0.01	-0.04	0.32	0.29	0.10
Bright, harsh fluorescent lighting can make me feel tense	0.49	-0.15	0.36	0.40	0.32
The quality of light in my workplace is irrelevant to my job satisfaction	0.21	-0.12	0.12	-0.03	-0.01
Natural daylight indoors improves my mood	0.09	-0.12	-0.27	0.01	0.19
Bright lights in grocers and pharmacies don't make me buy more	0.02	-0.08	-0.30	0.06	0.18
It makes no difference to me what kind of lighting is in a room	-0.48	0.16	-0.49	-0.39	-0.43
Soft, diffuse light is soothing	0.01	-0.10	0.25	-0.02	-0.24
Pregnant women should avoid exposure to fluorescent lighting	0.05	-0.31	0.36	0.35	0.48
My vision never becomes blurred when the lights are very bright	-0.10	-0.09	-0.35	-0.11	0.23
Glaring lights give me headaches	0.11	0.06	-0.09	-0.15	-0.10
I rarely use warm-coloured lighting to help me relax	0.02	-0.15	-0.16	0.00	0.26
Reading under dim light doesn't damage your vision	-0.39	0.54	0.08	-0.18	-0.40
Fluorescent lights are bad for your health	-0.04	0.39	0.12	-0.25	0.04

Table 7-20 (continued next page): Correlations between subject's change in perception scores between electric lighting and TDDs, and their agreement with various statements about their beliefs/attitudes towards lighting. Note: significant ($p<0.05$) correlations are highlighted green, while non-significant ones are grey. Marginal ($p<0.01$) correlations are highlighted blue.

	Lighting Quality	Glare Problems	Daylight Quality	Room Attractiveness	Illumination
If I want to create an intimate setting, I dim the lights	0.08	0.02	0.06	-0.05	-0.17
Bright light makes people talk louder	0.11	0.15	0.15	0.05	0.23
Fluorescent light seldom gives me a headache	-0.28	-0.03	-0.24	-0.31	-0.48
If a restaurant is very brightly lit, I will leave soon after I've finished eating	0.65	-0.12	0.16	0.48	0.58
Lack of sunlight in winter does not bother me	0.06	-0.17	0.23	0.17	-0.09
Incandescent lighting in a room helps me to pay attention to the speaker	0.49	-0.33	0.54	0.43	0.23
I do my best work in places that are lit using natural daylight	-0.24	-0.01	0.00	-0.06	-0.14
Humming noise from fluorescent lights usually does not distract me	0.19	0.16	-0.17	-0.10	-0.17
Bright lights rarely make me feel excited and full of anticipation	-0.19	0.22	-0.03	0.05	0.11

Table 7-20 cont.: Correlations between subject's change in perception scores between electric lighting and TDDs, and their agreement with various statements about their beliefs/attitudes towards lighting. Note: significant ($p<0.05$) correlations are highlighted green, while non-significant ones are grey. Marginal ($p<0.01$) correlations are highlighted blue.

Appendix G: Controls

G1: English as a second language

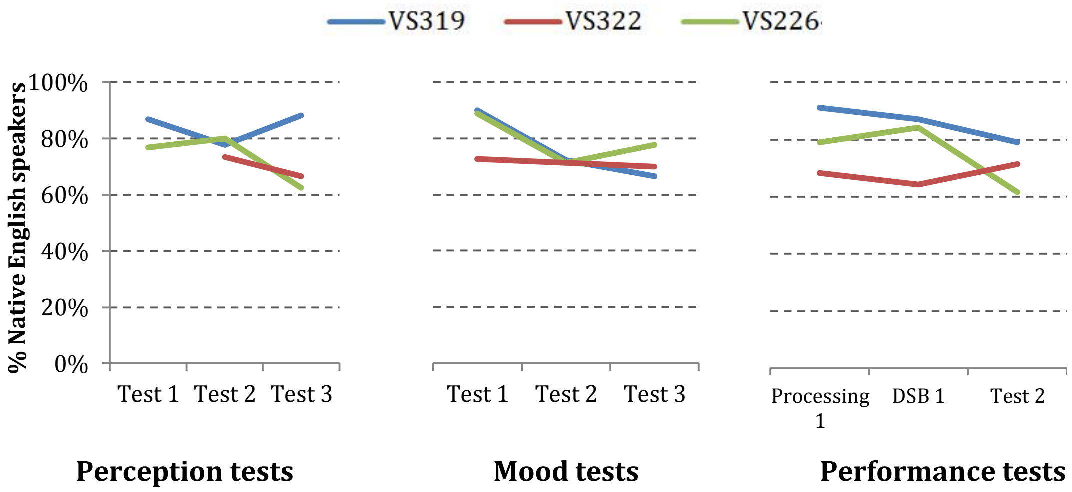


Figure 7-15: Proportion of sample that are native English speakers in the different rooms and tests

	ESOL: *n* = 14		Native English: *n* = 55				
	Mean	SD	Mean	SD	Mean difference	t	significance
Lighting Quality	2.42	0.57	2.29	0.73	0.13	0.74	ns
Glare Problems	1.20	0.90	1.46	1.17	-0.25	-0.88	ns
Daylight Quality	0.31	0.72	0.18	0.50	0.13	0.65	ns
Room Attractiveness	3.91	1.59	3.48	1.22	0.44	0.96	ns
Illumination	4.82	1.50	4.18	1.15	0.63	1.47	ns

Table 7-21: Perception results divided by native language using the average of all the groups under electric lighting

	Native English speaker			ESOL					
	n	Mean	SD	*n*	Mean	SD	Mean difference	t	significance
DSB	51	3.15	1.22	12	2.92	0.95	0.23	0.71	ns
Processing	48	14.74	3.32	10	14.36	2.41	0.38	0.42	ns

Table 7-22: Performance test results divided by native language using the results of the first round of testing under electric lighting

	Native English speaker			ESOL					
	n	Mean	SD	*n*	Mean	SD	Mean difference	t	significance
Δ Pleasure	62	-0.15	1.20	17	-0.36	1.16	0.21	0.67	ns
Δ Arousal	62	0.04	0.67	17	-0.07	0.79	0.11	0.52	ns

Table 7-23: Mood change results divided by native language using the results from the first two surveys under electric lighting

	Native English speaker			ESOL			Mean difference	t	significance
	n	Mean	SD	*n*	Mean	SD			
Sleepiness	79	5.06	1.32	19	5.20	1.75	0.15	0.34	ns

Table 7-24: Sleepiness results divided by native language using the results from the first half of the semester under electric lighting.

G2: Wearing glasses

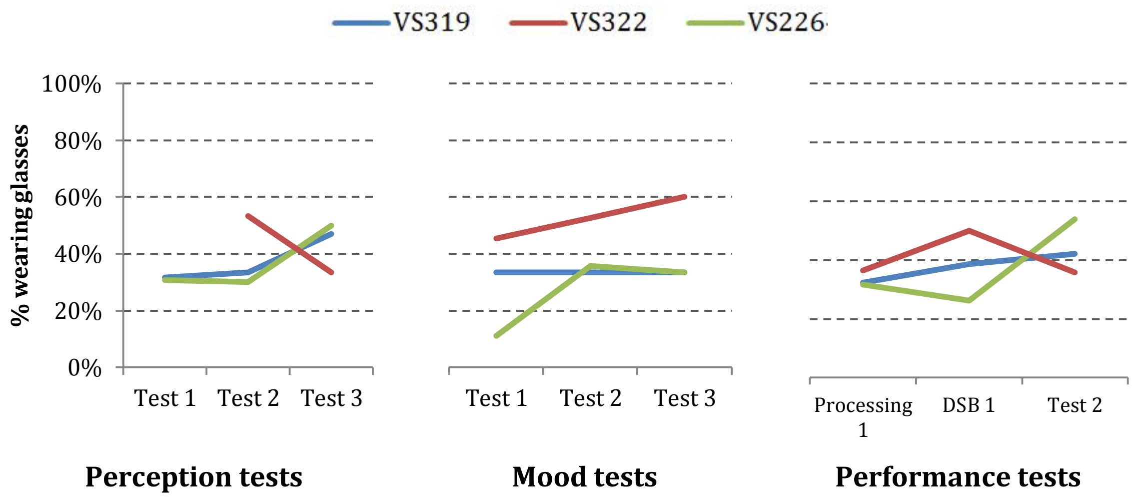


Figure 7-16: Proportion of the sample that wear glasses in the different rooms and tests

	Glasses: *n* = 25		No glasses: *n* = 44		Mean difference	t	significance
	Mean	SD	Mean	SD			
Lighting Quality	2.37	0.73	2.29	0.69	0.09	0.48	ns
Glare Problems	1.45	1.28	1.38	1.03	0.07	0.22	ns
Daylight Quality	0.25	0.66	0.17	0.48	0.08	0.52	ns
Room Attractiveness	4.00	1.34	3.76	0.96	0.25	0.81	ns

Table 7-25: Perception results divided by whether or not they wear glasses using the average of all the groups under electric lighting

	Wears glasses			Doesn't wear glasses			Mean difference	t	significance
	n	Mean	SD	*n*	Mean	SD			
DSB	24	2.81	1.21	39	3.28	1.12	-0.47	-1.54	ns
Processing	20	14.17	3.57	38	14.94	2.94	-0.77	-0.83	ns

Table 7-26: Digit Span Backwards and Processing Speed results divided by whether or not they wear glasses using results of the first round of testing under electric lighting

	Wears glasses			Doesn't wear glasses			Mean difference	t	significance
	n	Mean	SD	*n*	Mean	SD			
Δ Pleasure	26	-0.35	0.97	53	-0.12	1.28	-0.23	-0.88	ns
Δ Arousal	26	-0.03	0.48	53	0.04	0.78	-0.06	-0.45	ns

Table 7-27: Mood change results divided by whether or not they wear glasses using the results from the first two surveys under electric lighting

	Wears glasses			Doesn't wear glasses			Mean difference	t	significance
	n	Mean	SD	*n*	Mean	SD			
Sleepiness	31	5.08	1.44	67	5.09	1.40	-0.01	-0.03	ns

Table 7-28: Sleepiness results divided by whether or not they wear glasses using the results from the first half of the semester under electric lighting.

G3: Whether or not they've eaten breakfast/lunch

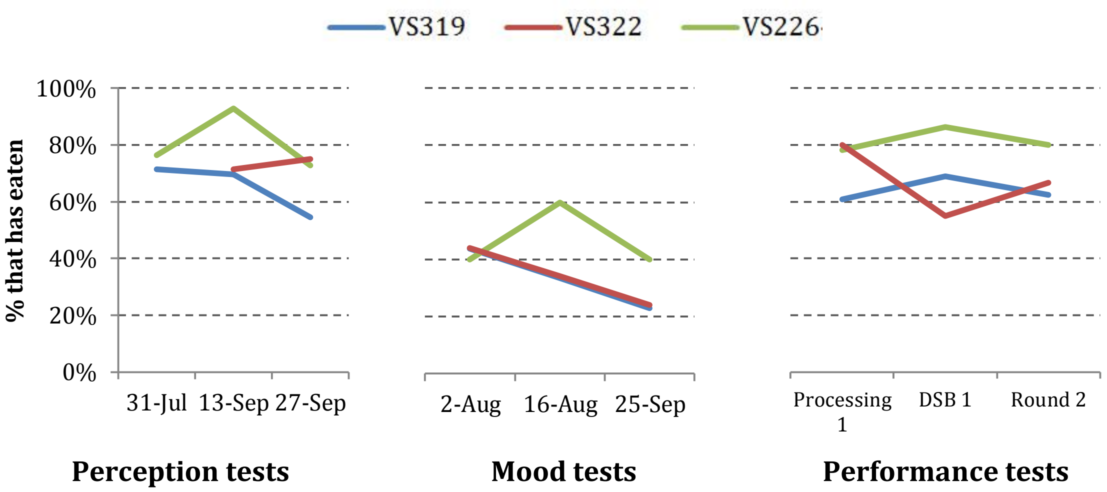


Figure 7-17: Proportion of the sample that had eaten lunch/breakfast beforehand broken up by room and test

	Has eaten: *n* = 63		Has not eaten: *n* = 24		Mean difference	t	significance
	Mean	SD	Mean	SD			
Lighting Quality	2.23	0.74	2.65	0.73	-0.42	-2.36	p<0.05
Glare Problems	1.33	1.17	1.77	1.27	-0.44	-1.47	ns
Daylight Quality	0.27	0.68	0.58	0.93	-0.31	-1.51	ns
Attractiveness	3.40	1.32	4.25	1.30	-0.85	-2.71	p<0.05
Illumination	4.20	1.32	4.64	1.30	-0.44	-1.41	ns

Table 7-29: Perception results divided by whether or not they had eaten beforehand using the average of all the groups under electric lighting

	Has eaten			Has not eaten			Mean		
	n	Mean	SD	*n*	Mean	SD	difference	t	significance
DSB	60	3.12	1.27	26	3.33	1.17	-0.21	-0.75	ns
Processing	53	14.77	2.84	19	15.26	3.36	-0.32	-0.36	ns

Table 7-30: Performance test results divided by whether or not they had eaten beforehand using the results of the first round of testing under electric lighting

	Has eaten			Has not eaten			Mean		
	n	Mean	SD	*n*	Mean	SD	difference	t	significance
Δ Pleasure	71	-0.18	1.27	27	-0.04	1.04	-0.14	-0.57	ns
Δ Arousal	71	0.18	0.93	27	-0.10	0.60	0.29	1.81	p<0.1

Table 7-31: Mood change results divided by whether or not they had eaten beforehand using the results from the first two surveys under electric lighting

	Has eaten			Has not eaten			Mean		
	n	Mean	SD	*n*	Mean	SD	difference	t	significance
Sleepiness	58	4.66	1.75	29	5.59	1.88	-0.93	-2.23	p<0.05

Table 7-32: Sleepiness results divided by whether or not they had eaten beforehand using the results from one day under electric lighting.

G4: Whether or not they've drunk coffee

	Has drunk coffee: *n* = 30		Has not drunk coffee: *n* = 57		Mean		
	Mean	SD	Mean	SD	difference	t	significance
Lighting Quality	2.17	0.85	2.44	0.69	-0.27	-1.49	ns
Glare Problems	1.62	1.23	1.37	1.19	0.25	0.90	ns
Daylight Quality	0.20	0.48	0.44	0.87	-0.24	-1.65	p<0.1
Attractiveness	3.35	1.32	3.78	1.37	-0.43	-1.42	ns
Illumination	3.90	1.20	4.54	1.34	-0.64	-2.26	p<0.05

Table 7-33: Perception results divided by whether or not they had drunk coffee beforehand using the average of all the groups under electric lighting

	Has drunk coffee			Has not drunk coffee			Mean		
	n	Mean	SD	*n*	Mean	SD	difference	t	significance
DSB	21	3.64	1.04	65	3.03	1.26	0.61	2.22	p<0.05
Processing	22	14.77	2.84	50	15.15	3.43	-0.38	-0.49	ns

Table 7-34: Performance test results divided by whether or not they had drunk coffee beforehand using the results of the first round of testing under electric lighting

	Has drunk coffee			Has not drunk coffee			Mean difference	t	significance
	n	Mean	SD	*n*	Mean	SD			
Δ Pleasure	22	-0.02	1.11	76	-0.18	1.24	0.16	0.59	ns
Δ Arousal	22	-0.03	0.80	76	0.14	0.88	-0.17	-0.88	ns

Table 7-35: Mood change results divided by whether or not they had drunk coffee beforehand using the results from the first two surveys under electric lighting

	Has drunk coffee			Has not drunk coffee			Mean difference	t	significance
	n	Mean	SD	*n*	Mean	SD			
Sleepiness	21	4.52	1.57	65	5.38	1.82	-0.86	-2.10	p<0.05

Table 7-36: Sleepiness results divided by whether or not they had drunk coffee beforehand using the results from one day under electric lighting.