Validation of the 4pHAT

Dissertation submitted in partial fulfilment of the

requirements for the Degree of

Master of Science in Applied Psychology

at the University of Canterbury

Jason Hunt

University of Canterbury

2019

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Table of Contents

List of Tables

List of Figures

Acknowledgements

Firstly, I would like to thank the time given by the participants, and their acquaintances, to the present research. Without them, this research would not have been possible.

Thank you to my research supervisor, Chris Burt for his knowledge and guidance on this project. His specific knowledge and expertise over the tools in question was invaluable. Thank you also to Joana Kuntz for being an ever supportive, encouraging and sarcastic associate supervisor.

A Special thank you to the technical team behind the research. To Kezia and Richard from Talegent for their supportive partnership, Ben in the Psychology workshop for his work in the hazard lab and Gerard for fixing all my computer problems.

Thank you to the Tutor and RA teams at College House and University Hall over the past two years, you have been a significant part of this journey and I thank you all for your support and patience.

Finally, and fittingly, thank you to my APSY classmates over the last year. We survived the high attrition and made it to the end, poorer in one sense but much richer in another. Thank you for the constant unrelated discussions to procrastinate with and for being there as a sounding board when required. It could not have been done without you all and I look forward to following your paths into the future.

Abstract

Both in New Zealand and abroad, the rate of workplace accidents is extremely high and leads to significant human and financial loss. While measurement alone will not solve the issue, organisations measuring health and safety-related behaviours and abilities can provide an important step in reducing potential adverse events. Unfortunately, the existing tools to measure factors are susceptible to issues around self-report or do not measure appropriate behaviours predictive of accident outcomes. As hazard recognition is a leading cause in workplace accidents, the Hazard Awareness Test (HAT) was developed as a gamified, objective tool to measure job candidates hazard recognition abilities. Past validation of the HAT showed promising findings, but the validation methods were concurrent rather than predictive. The present study had two main aims; to further validate the HAT as a criterion valid tool using a predictive validation method and to validate a shorter, more commercially viable version of the tool, the 4pHAT. Seventy participants completed the HAT/4pHAT and a lab-based hazard recognition task one week later. In addressing the criterion-related validity of the measures, the results indicated that the HAT and the 4pHAT are effective measures of hazard recognition ability. The implications of these results in an applied context are discussed.

Introduction

Overview

The introduction begins by examining the current state of organisational health and safety both in New Zealand and abroad, addressing both the human and financial cost of workplace accidents. The current New Zealand health and safety legislation is reviewed and its impact on the workplace is addressed. Insight is then given to the theoretical rationale behind the causes of accidents at both an organisational and at the individual level. The role that hazard recognition plays in the accident process is then examined, as well as how it is developed as an ability through experience and gained safety knowledge. Given many accidents and incidents are associated with human behaviour it is useful for organizations to screen individuals for safety-related abilities at the time of recruitment. Thus, how safety ability and behaviours can be measured during the employee selection process using tools that are currently available in the market is reviewed. Potential pitfalls of these measures, as well as the limitations of self-report assessment such as interviews, are addressed, with a discussion on how these issues could be solved through the development of more objective measures. Two such measures of hazard recognition ability are described – the HAT and the 4pHAT, along with the gamification principles used for their development. The introduction then turns to the focus of this work being an investigation of the criterion-related validity of the HAT and 4pHAT, and the method used to validate them is discussed, along with an examination of the criterion measures. The introduction concludes with several predictions.

Health and Safety in New Zealand and Abroad

According to the International Labour Office (ILO) (2018), more than 2.8 million deaths per year can be attributed to occupational accidents or work-related diseases. A further 374 million non-fatal work-related injuries and illnesses occur each year. As well as the significant human loss, the economic cost of poor health and safety practices leading to death and injury is estimated to be close to 4% of global gross domestic profit yearly. Even more troubling is potentially up to 80% of workplace incidents are going unreported, suggesting the actual impact could be significantly greater (Probst, Barbaranelli & Petitta, 2013).

New Zealand doesn't fare any better. Based on ACC data and reports to WorkSafe New Zealand, 72 workplace fatalities occurred in New Zealand across all industries in the 2017 calendar year, as well as close to 3000 serious workplace injuries or illnesses (WorkSafe, 2018). In that same year, 231,100 work-related claims were made to ACC, with the number increasing since 2012 (ACC, 2017). It was estimated that in 2010, work-related incidents cost the country approximately $3.5 billion, or around 2% of GDP. Though depending on how these costs are measured, the estimate is closer to $20 billion (Independent Taskforce on Workplace Health and Safety, 2013). The nature of our economy in New Zealand lends itself to high-risk industries such as agriculture, inflating our accident rates. Manufacturing, construction, agriculture, forestry and fishing make up more than 50% of ACC claims each year as well as having the highest rates of claims, ranging from 24 to 32 claims per 1000 full-time workers (Independent Taskforce on Workplace Health and Safety, 2013).

New Zealand Health and Safety Legislation

2016 saw a major shakeup in the health and safety legislation of New Zealand. Following the independent task force's review of workplace health and safety in 2013, the Health and Safety at Work Act 2015 (HSWA) was developed, alongside the establishment of WorkSafe New Zealand (WorkSafe New Zealand, 2017a). The main goal of the new Act was to afford workers "the highest level of protection from workplace health and safety risks, as is reasonable" (WorkSafe New Zealand, 2017a), through a change from the previous recording and monitoring focus of health and safety practices, to a proactive system that attempts to manage risks before they become problematic. Among the changes was the shift of primary responsibility to organisations. While the Act is clear about health and safety being everyone's responsibility to some degree, the onus on organisations to ensure safe working environments, mitigate risks where possible, and to carry out due diligence, shone a much-needed light on the importance of health and safety. Significant penalties were put in place by the HSW Act to create liability to encourage organisations to follow the new legislation. If an employee is exposed to the risk of injury, illness or death, without a reasoned excuse and the employer did not meet their responsibilities outlines in the HSWA, they are liable for a fine of up to $600,000 or up to 5 years in prison **(**Section 47, Health and Safety at Work Act, 2015). The penalties are even more severe for the organisation, with a fine of up to $3 million able to be handed down.

The employee themselves also has direct reasonability's under the HSWA. Firstly, they must be responsible for their own health and safety. For example, if they believe they are being exposed to risk, they may cease working, so the situation can be remedied (WorkSafe New Zealand, 2017b). They must also ensure their actions do not put others at risk and must comply with reasonable health and safety policies, procedures or instructions given by the employer (WorkSafe New Zealand, 2017b). The HSWA encourages communication and has provisions for worker engagement in the form of health and safety representatives and committees. While workers are under no legal obligation to take part in this communication process it is hoped the HSWA creates the ability and willingness for this to happen.

Workplace Health and Safety

Within the workplace, safety is defined as the state of operations where the risk of damage, loss or injury is minimised or eliminated (Young, 2012). Workplace health and safety comprises of all activities within the workplace and does not only apply to high-risk sectors such as construction and agriculture, with seemingly low-risk environments such as offices also included in the definition (Lamm, 2009). Organisations now, due to the legal liability placed on them by the HSWA (2015), must invest time and resources into ensuring their workplaces are as safe as is practical. In the past, investment into health and safety was limited due to no clear return on investment coming from it (Maudgalya, Genaidy & Shell, 2008). In the current climate, the cost of poor health and safety practices leading to unsafe acts in the workplace can have significant legal penalties, but further costs are also prevalent. Both direct costs, such as loss of employee time or plant damage, and indirect cost, including recruiting new employees to replace injured ones, or damage to reputation, can have a substantial impact financial impact on the organisation (Neal & Griffin, 2006). For organisations to fully take charge of their health and safety practices and reduce their accidents rates, an understanding of how and why accidents occur can be highly beneficial. Research has developed a number of models of accident causation which are examined in the next section.

Accident Models

Accidents will most often have several factors that influence their outcome. These factors arise across all levels of the organisation, from the policies put in place by the organisation, though to the behaviour of supervisors and the actions of the workers on the ground (Shappell & Weigmann, 2000). The root causes of these accidents generally come from pre-existing issues within the organisation's safety systems (Reason, 1990). Shappell and Weigmann's (2000) Human Factors Analysis and Classification System (HFACS) clearly lays out the potential influences on an accident occurring. The model identifies four key stages across an organisation where a failure needs to occur, for an accident to result. Firstly, issues surrounding the top level of the organisation, such as poor selection procedures, failed chain of command or poor safety programs create the first step for an incident to occur. Next, failures in supervision and leadership on the ground, such as a lack of training or inappropriate directions create the next step. Then, any adverse physical, mental or psychological conditions that an individual is under, leads to the final step in which an individual either has an error in their performance (unintentional) or internationally violates correct behaviours in the situation. While organisations have the obligation to create an environment that is as safe as is reasonable, it is accepted that things within a large system may be missed and therefore leave the system vulnerable to failures. This is where the abilities of individuals have an effect. Unintentional failures on the part of the individual such as the failure to identify a risk due to a lack of hazard knowledge or perceptual abilities can act as the final catalyst in a system to cause an accident (Reason, 1990).

Ramsey's (1985) Accident Sequence Model seeks to explain the process individuals go through from when they are exposed to a hazardous situation to a potential accident occurring. After exposure to a hazard, they must first use their sensory abilities to physically perceive the hazard. Next, utilising past knowledge, they must realise that what they have perceived is hazardous and therefore dangerous to them. They then need to decide to avoid the hazard and finally need to have the physical ability to avoid the hazard. If any of these four steps are violated, it will likely lead to an accident occurring. Key to this process is hazard recognition.

Hazard Recognition

Hazard recognition is the first step in the safety management process (Perlman, Sacks & Barak, 2014), and is concerned with the first two steps in Ramsey's (1985) model; the sensory perception of a stimulus and the knowledge that is it a hazard. Failure of hazard recognition, by the individual in this step, not only increases the chance of an acute failure but also increases the risk of future accidents (Carter & Smith, 2006). If workers are failing to identify hazards in their working environments, and therefore not communicating these issues, it is not possible for organisations to rectify these issues. This also limits the ability of organisations in New Zealand to meet their obligations under the HSWA to mitigate the risk of the hazards that are present **(**Health and Safety at Work Act, 2015).

Health and safety statistics reflect certain industries as having greater accident rates than others. It seems reasonable to assume that high risk working environments, such as construction, forestry and agriculture would have a larger number of accidents compared to office work, given the dynamic nature of the working environment and the number of potential hazards present. The new health and safety legislation in New Zealand should have an impact on the accident rate, from an organisational perspective. Conditions for unsafe acts to occurs should be greatly reduced by a larger focus on reducing risk in the first place. This now means, more than ever, that a focus on individuals safety performance is paramount to reducing the number of workplace accidents further. Haslam, et, al. (2005) conducted a review of 100 incident reports of accidents occurring in the construction industry. Seventy percent of the reports highlight individual workers or work teams as a key factor in accidents, similar to the 70 – 80% figure sighting ‘human error’ as a determining factor in accident reviews (Rasmussen, 1997). Of significance to the present study was that in 42% of the cases investigated by Haslam, et, al. (2005), inadequate hazard recognition could be linked to the cause of the accident.

Two key factors make up the hazard recognition process; detection and knowledge. Kowalski-Trakofler and Barrett (2003) pose a five-step model of hazard recognition: Detection of sensory cues – Attentional Selection – Recognition of Hazard – Confirmation of Hazard – Appropriate Response. These five steps can be grouped into two core factors. For an individual to detect a hazard they must sense the hazard and give attention to it (Jeelani, Albert, Azevedo & Jaselskis, 2016). To conclude that what has been sensed is a hazard, an individual utilises their past knowledge of hazards to perceive and confirm the hazard. The final step proposed by Kowalski-Trakofler and Barrett (2003) goes beyond the recognition of a hazard and is more concerned with the behavioural response once a hazard has been identified.

Bahn (2013) conducted a workshop assessing the hazard recognition abilities of 77 miners in Australia. Workers were given clear examples of hazards across four different typologies (obvious, trivial, emerging and hidden hazards), they were then asked to identify the hazards encountered in their own work areas. The study found that on average, only 46% of the hazards were identified. While those with less experience underground in the mines (3 years or less), performed worse than the rest, it was concluded that more training was needed for all members of the organisation, with even experienced workers and supervisors performing poorly. Perlman, Sacks and Barak (2014) conducted a study assessing the hazard recognition abilities of civil engineering students, site superintendents and safety directors. They were made to identify hazards that would be seen in a typical construction project via one of two methods. The first was identifying hazards presented on a photograph of construction sites or on construction drawings. The second method utilised a three-sided virtual reality construction site in which individuals were to identify 48 different hazards. The virtual reality method aimed to replicate a real site without exposing participants to any undue risk. This method aligns with the 'Hazard Lab' utilised in the present study, with further discussion on this in a later section. Perlman, Sacks and Barak (2014) found that the percentage of hazards correctly identified by participants ranged from 27% to 44%, with no significant differences being identified between the three different subject groups. This suggested that experience level on the job had no effect on individual hazard recognition abilities.

The danger of a failure of hazard recognition can be two-fold. Firstly, it increases the potential of individuals being involved in an accident, as they are not picking up potential risks. Secondly, the ability that the management of an organisation has to reduce further incidents is significantly reduced. If workers are not identifying hazards in their workplace, it is not possible for said hazards to be removed or mitigated, potentially leading to more accidents (Albert, Hallowell, Skaggs & Kleiner, 2017; Namian, Albert, Zuluaga & Jaselskis, 2016).

Hazard recognition is an ability that can be developed in an individual over time through training and experience (Albert, Hallowell & Kleiner, 2013). Carter and Smith (2006) conducted a study assessing the hazard identification of three construction sites in the United Kingdom. They reviewed method statements from the sites, that included a detailed report on the hazards identified on the site and compared them with the actual number of hazards on the site (identified by the researchers). Hazard identification ranged from 67% to 90% across the three sites, with Cater and Smith concluding these results as being 'far from ideal'. While the study did not look at hazard recognition at the individual level across the site, those completing the site method statements had to utilizes their own hazard recognition skills. What is of most relevance to the present study was their proposal of a lack of knowledge being a barrier to improving hazard recognition. This knowledge would be gained either through past experience in similar situations or through specific training to increase their abilities (Perlman at, al., 2014). Albert et, al. (2013) put workers through a program that taught them how to decompose basic job tasks to; identify hazards, identify the role they played in communicating said hazard and reviewed the situation to notice any changes that needed to be made in their hazard recognition performance. After the program, workers showed a 31% increase in hazard recognition on construction site photos. The process increased the safety knowledge of all the workers, becoming aware of what was and wasn't a hazard, and made them more attentive to perceiving hazards in their environments. As hazard recognition is able to be learned and increased, it is, therefore, going to be on a spectrum for the general population. This means that individuals will vary in their hazard recognition abilities, so it seems obvious then to measure this ability in prospective employees. In this selection setting, it can show the current levels of hazard recognition ability an individual has, and therefore whether they will be suitable for the role.

Safety in Recruitment and Selection

The number of employees within an organisation that fail to behave in a safe manner, or have poor safety abilities, has a direct relationship with the number of incidents and accidents within said organisation (Neal & Griffin, 2006). Organisations can reduce the number of employees with poor safety behaviour, and consequently reduce the number of accidents, by utilising the recruitment and selection process to predict the safety behaviours of potential new employees.

In the past, this was assessed through data pertaining to candidates past accident history. However, this is a problem with the assumption that past accident involvement will predict future accident involvement. Laughery and Vaubel (1989) reviewed a database of 12786 accidents for 2259 different individuals, collected from a petrochemical complex over an 11-year period. They found significant positive correlations between individuals past accident experience and their current safety behaviours. Similar results were found by Kouabnan's (2002) study of the road accidents of 553 individuals. They found that drivers who had previously been involved in a vehicle accident were less likely to engage in risky behaviours, such as speeding, and more likely to be more careful and cautious drivers (m =

82.42 versus 85.96 on the risk-taking index). While the relationship between past accidents and safety behaviour is clearly outlined by these studies, it still poses a significant issue. This method does not give employees an idea of the safety behaviours of those who have never been involved in an accident. More predictive approaches to identify safety behaviours would be needed for these individuals. These predictive approaches could be in the form of competency-based application blanks, or through structured interviews assessing specific job-related safety behaviours (Levashina, Hartwell, Morgenson, & Campion, 2014: Wood & Payne, 1998).

Interviews and Self Report Limitations

One major limitation that can arise from an application black or interview method of obtaining safety behaviour information for candidates, stems from issues with self-reported data. Self-report is prone to several issues, the most prevalent of which is social desirability bias (Arnold & Feldman, 1981). Individuals want to present themselves in the most favourable light that they can, in order to appear better than they are (Podsakoff, & Organ, 1986). Imagine a question in which candidates for a construction role are asked about their previous safety behaviour in the workplace. It may be framed as *"Please describe the behaviours you conducted daily while on site that helped to keep you and your colleagues safe".* The candidate is inevitable going to respond in a way that they believe is what the interviewer wants to hear with responses such as "wearing all appropriate safety equipment" or "communicating all hazards on a construction site to my supervisor". Directly related to social desirability bias in interviews is individual's tendency to fake. Across six studies, with 1346 participants, Levashina and Campion (2007) showed high rates of faking and lying during employment interviews. Faking behaviours were grouped in three separate categories; image creation (creating the image of a good candidate), image protection (protecting a favourable image) and ingratiation (attempting to gaining favourable perception from the interviewer). They found that in more than 90% of cases, candidates displayed some level of faking behaviours in employment interviews. The tendency for candidates to fake by giving socially desirable responses reduces the validity of the interview as an effective method to gain information on a candidate's safety behaviours. To address this, a number of psychometric tools have been developed.

Psychometrics in Health and Safety Selection

A 2010 review by Barrett identified 15 different tools currently available in the market to assess the different attributes of safety for both current and prospective employees. Table 1 shows a list of these tools with their associated publisher. These tools measure a range of factors such as safety knowledge, compliance, participation, motivation and adherence to rules in an attempt to predict safety behaviour in the workplace (Barret, 2010). All the measures listed below contain at least one personality-based factor being measured, with the majority only using personality measures as predictors of safety. It is evident that organisations are interested in assessing some health and safety component of prospective employees at the selection stage, given the number of measures that have been developed. However, little attention has been given to identifying and accurately measuring the factors that determine an individual's safety outcomes (DeJoy, et, al., 2004). A meta-analysis of some 90 studies, conducted by Christian et, al. (2009), investigated the relationship between personality factors found across the studies, with both 'safety performance' and 'actual safety outcomes'. The safety performance factor was an aggregate of safety compliance (Safety-related behaviours required by the organization) and safety participation (voluntary behaviours that had no effect on personal safety but contributed to safety in the greater organisational context). Actual safety outcomes referred to accidents and injuries experienced by participants in the review studies. The results showed that while the average correlations between the personality measures and the ‘safety performance’ measure were .29, it fell to .19 when measured against ‘actual safety outcomes’. When utilising a personality passed test to predict an area of job performance, such as safety performance, Schmidt and Hunter’s (1998) meta-analysis of 85 years of validation research, identified that a correlation value of .31 was required for a measure to be of significant validity. It is clear from these findings that the current crop of measures likely fails to accurately predict the safety outcomes of candidates and therefore like does not have the potential to reduce workplace accidents.

Psychometric tools are also not exempt from the self-report and social desirability issues highlight by interviews. Van de Mortel’s (2008) conducted a review of 31 studies with self-report measures that also included social desirability scales. The results showed that in 43% of the studies, the effect of social desirability had a significant impact on the results, to the degree that Van de Mortel concluded the results of such studies could be ‘flawed’. In only 10% of the reviewed studies did the researcher's address or control for social desirability after it was identified in the results. King and Bruner (2000) review of nearly 20 years’ worth of published research regarding self-report scale design and social desirability bias, showed a similar trend. Their review noted that in studies involving scale construction, more often than not, scales were not modified in any way, even when significant relationships were found between the scale constructs and social desirability scale measures, such as the 33-item Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960), the most widely used measure of social desirability.

Both King and Bruner (2000) and Van de Mortel’s (2008) suggest the researchers have the ability to address the issues of social desirability in their scales, but they are failing to do so. Researchers designing future scales need to make a concerted effort to address such issues or go about methods that reduce these issues altogether. There is a clear need for tools to be developed that address both the limitations of interviews for obtaining health and safety behaviour information, that reduce the effect of social desirability bias and act as better predictors of actual safety abilities and outcomes.

Table 1. *Currently available Occupational Health and Safety Measures*

Commercial Product	**Publisher**
Employee Reliability Inventory	Bay State Psychological Associates Inc.
Health and Safety Indicator (HSI)	OPRA Consulting Group
Personnel Reaction Blank	IPAT Inc.
Onetest Work Safety Assessment (OWSA)	OneTest Pty Ltd.
Orion Pre Employment System PE3-SAFE	Orion System Inc.
Situational Safety Awareness Test	Psyfactors Pty Ltd.
Work Safety Assessment	Psych Press
Risk Type Compass	Psychological Consulting Ltd. (PCL)
Work Attitude Inventory (WAI)	Psytech International Ltd.
RMP Safety Inventory	RightPeople
Workplace Safety Solution Test	SHL Plc.
Safety Attitude Survey	Synergy Safety Systems
Employee Safety Inventory (ESI)	Vangent (Pearson) Inc.
Personnel Selection Inventory (PSI)	Vangent (Pearson) Inc.
Hogan Safe System	Hogan Assessment Systems Inc.

Note: Adapted from Barrett's (2010) review of commercial products associated with the psychological assessment of safety attributes within prospective employees.

Hazard Awareness Test

One such measure that was developed to address some of the pitfalls of the current measures in the health and safety space, as mentioned above, and to reduce accidents caused by a lack of hazard recognition is the Hazard Awareness Test (HAT). The development of the HAT was conducted at the University of Canterbury, led by Associate Professor Christopher Burt. The HAT was designed, with gamification principles in mind, as a test to measure an individual's hazard awareness ability. The HAT utilises a 'spot the difference' approach in which individuals are presented with a side-by-side image pair with ten differences between the left-hand image and right-hand image. Like other 'spot the difference' puzzles the differences included changes in the colour of specific items, the removal or addition of items and subtle manipulation in the positioning of items. The HAT consists of 10 image pairs depicting scenes from five different life domains, including; workshop safety, office safety, outdoor work safety, water safety and home-life safety. Each scene includes five safety-related differences (e.g. worker wearing then not wearing a hard hat) and five neutral differences (e.g. shirt colour changing from green to blue). Developing scenes across a range of settings was done under the premise that safety-conscious individuals (those that should score highly on the HAT) will have learnt about hazards from several different life domains (Burt, 2017).

Gamification

The HAT was developed using gamification principles, and as such provides a solution for reducing the biases associated with many, if not all other measures of safety. Gamification is the use of game design elements (aspects of game-based technology rather than fully-fledged games), not for the purpose of play or recreation, but in non-game contexts (education, measurement, training, assessment, etc.) (Deterding, Dixon, Khaled, & Nacke, 2011). Game design elements may consist of leaderboards, scores, competition, turn taking, time limits, levels and rewards. Incorporating these elements into regular tasks attempts to increase motivation on the task, engagement, and overall increases the user's experience (Buckley & Doyle, 2016; Hamari, Koivisto & Sarsa, 2014). This could be as simple as adding a competitive element (game design element) to an in-class quiz, having children competing against each other to accumulate points to win a prize. At the more extreme end, it may be a firefighter using virtual reality goggles to immerse themselves in a realistic training environment.

Gamification has been a growing trend around the world in the last 15 years. Education is the largest adopter of gamification which is reflected in a large proportion of the applied research being conducted in this field (Seaborn, & Fels, 2015). It is however also used widely across many industries including marketing, health and wellbeing, social media, crowdsourcing and business (Huotari & Hamari, 2012; Kumar, 2013; Seaborn, & Fels, 2015).

Gamified tools are being used across a wide range of organisational spaces. Given the massive uptake in the education sector to increase learning, organisations have adopted a similar approach to train their personnel. Electronic Arts (EA), one of the largest video game companies in the world, have created EA University, an internal training program that utilises their vast experience in video game development to provide employees with a dynamic environment to develop new work relevant skills (Lowman, 2016). Outside of the training space, gamified tools are being used for assessment of personality, intelligence and for specific job-related skills. HR Avatar, a major creator and supplier of pre-employment assessments around the world, utilise gamified tools using interactive cartoons to simulate workplace environments for customer service and security roles (Chamorro-Premuzic, Winsborough, Sherman, & Hogan, 2016).

A potential benefit of utilising gamification principles in selection tools is the ability to take the objective nature of games and apply in an assessment situation (Buckley & Doyle, 2016). One key gamification element is having objective, specific rules and the competitive outcome. The rules in which the gamified tool is developed predetermine the actions that can be taken by the test taker, meaning only that response is possible, mitigating any subjectivity in responses (Buckley & Doyle, 2016). While other tools could be created in this manner, the principles of gamification aid in the development. Further to removing subjectivity, the gamified nature of a tool can have a positive impact in the measurement accuracy of a tool through improved performance (Attali & Arieli-Attali, 2015). Individuals become more immersed in the assessment which in turn increases concentration on the task, motivation, enjoyment and ultimately task performance (Csikszentmihalyi, 1991; Eickhoff, Harris, de Vries & Srinivasan, 2012).

While the literature is limited on the empirical effectiveness of gamification as a measurement tool in the recruitment and selective space, its effectiveness in other areas suggests gamification as a highly useful metric, providing great potential for utilisation in recruitment and selection. A review by Looyestyn et, al. (2017) investigated the relationship with gamification and increased engagement with online programs, such as single surveys or education courses completed over an extended length of time. Their results showed that in 80% of studies, gamification had a significant positive relationship with increased engagement. Hamari, Koivisto, and Sarsa (2014) conducted a more general review of gamification. In answering the question "Does gamification work?" they analysed 24 empirical studies using psychological outcomes (motivation, attitude, enjoyment), behaviour outcomes (engagement) and motivational affordances (points systems, leaderboards, achievement badges) as criterion measures. In 63% of the studies, positive results were found supporting that gamification does work. Through a review of gamification literature around use as an assessment tool in an educational setting, Menezes and De Bortolli (2016) concluded gamified tools were able to provide students with a valid form of educational assessment. They believed the incorporation of evaluative measures in a game setting provided greater integration between the content being assessed and the evaluation itself.

Previous HAT Studies

The initial testing of the HAT was conducted at the University of Canterbury through the dissertations of Hill (2012) and Shaw (2012). Shaw (2012) conducted a study using the HAT on 39 construction works. She measured total differences found, safety differences found, and safety order scores from the HAT against both self-report and supervisor ratings of Safety Participation, Compliance, Motivation, Knowledge, Voicing Co-worker caring, and Voicing Behaviours. Significant correlations were found between *total differences found* in a subset of five HAT puzzles (Falls, Forest, Office, Workshop and Workshop Gear) and Neal, Griffin and Hart's (2000) 4 item scale measuring safety motivation. The forest, workshop and office puzzles also showed significant correlations between *total safety differences found* and the safety motivation measure. These results suggest that safety motivated employees put greater care into finding safety differences in the puzzles, however, it is unclear whether this would translate to real-world safety behaviours. A significant correlation was also found between the *safety order* (finding safety-related differences before neutral differences) and safety knowledge, suggesting that individuals with more safety knowledge were potentially looking for safety-related differences first or they were more apparent to them. It should be noted that no significant correlations were found between employees' self-report ratings of safety behaviours and that of their supervisors. While significant correlations were found between the HAT and the safety motivation, suggesting the validity of the HAT, one must question the construct validity of the criterion (as it does not explicitly measure hazard recognition) and the reliability of using self-report data, given the previously discussed issues with self-report.

Concurrently to Shaw (2012), Hill (2012) tested the HAT on 60 undergraduate students. She investigated the relationship between HAT scores and the reported number of workplace incidents and injuries experienced by participants. Correlations showed participants that reported fewer workplace incidents performed better on the HAT. This result suggested criterion-related validity on the basis that a smaller number of reported accidents meant greater hazard recognition abilities. The relationship between the HAT and safety was further investigated via conducting a between groups analysis. Participants were split into an *accident group* signifying that had experience at least one workplace accident, and a *non-accident group.* Results indicated s significant difference between the mean of the two groups with the *non-accident group* identifying, on average a greater number of safety differences on the HAT. As with Shaw (2012), the self-report nature of the criterion measure must be taken as a potentially poor criterion measure, questioning the criterion-related validity evidence of the study.

Thomas (2016) conducted a study measuring the construct validity of the HAT. The study split the 90 participants across three quasi-groups. Group one contained tertiary students who made a low health and safety (HSE) expertise group. The second group made up of employees from high-risk workplaces, such as construction, manufacturing, agriculture, made up the medium. The final group was a high HSE group made up of health and safety managers. Two key sets of findings were found in the study. The first showed a significant difference between the low (43.63) and the medium (46.12) and high (46.56) groups with a greater number of safety differences found on the HAT for the latter. Time taken to complete the HAT was not significantly different between the groups suggesting no time-related influence. The second set of findings, which suggested construct validity of the HAT, showed significant correlations between safety differences found on the HAT (for all participants) and total number of jobs held ($r=0.23$, $p=.05$), number of jobs with associated safety risk ($r=0.25$, $p=.05$), tenure in current job position ($r=0.22$, $p=.05$) and hours of health and safety training education($r=0.20$, $p=.05$). The rationale that these findings support suggests that through work experience, safety education and health and safety training, individuals learn or acquire greater hazard recognition abilities.

Burt (2017) combined all previous HAT studies to deliver a number of key finds. The general findings of the four previous studies (Shaw, 2012; Hill, 2012; Thomas: 2016; Burt & Adams, 2016) with 249 individuals ranging from last year high school students, though tertiary students, workers, and health and safety professionals are summarised below:

- All four studies provide some support for the HAT as a measure of hazard awareness ability. With scores on the HAT being positively associated with higher levels of individual workplace safety.
- Of potential control variables, age has been found to have a positive correlation with HAT scores (r =.18, p <.01, n=249). This is likely influenced by experience and learning that comes with age. This is further supported under the theoretical notion that through work experience, safety education, and health and safety training, individuals learn or acquire greater hazard awareness abilities.
- No evidence supports any gender difference in HAT scores.

One of the major findings, and a key to the success of the HAT as an objective measure of performance, is that scores are not able to be generated though bias responding or faking. Removing any aspect of self-reported responses from the measure means individuals are less able to manipulate responses (Cascio & Aguinis, 2011). The simplicity of the HAT means individuals either see a difference and click on it, or they don’t. There is no way to perform better than the individual is capable of, due to only having 10 chances (clicks) to identify the 10 differences.

The combined results of the previous four studies all formed the basis of the cut-off score for the use of the HAT in a commercial selection space. A score of 45 or greater would have individuals labelled as low risk. This value is based on Hill (2012) indicating that 83% of participants that scored 45 or greater had no history of workplace accidents requiring medical treatment. Burt and Adams (2016) had similar finds with 74% of participants falling into the same category.

4pHAT

Analysis by Burt (2017) suggested a shorter version of the HAT, using a subset including four of the 10 puzzles pairs, may be viable as a standalone tool called the 4pHAT. Burt (2017) examined scores obtained on the 4pHAT subset puzzles from all four of the previous studies (i.e. Shaw, 2012; Hill, 2012; Thomas: 2016; Burt & Adams, 2016). The results showed large significant correlations (p = .01) ranging between 0.80 and 0.86 across the 4 subset puzzle pairs. These results suggested that the hazard recognition ability being measured was evident and constant across all of the puzzles and that the variance captured was the same across the images. It was therefore assumed that little to no psychometric value was lost by reducing the full 10 puzzle pair version of the HAT down to the shorter 4pHAT version with only four. This assumption is tested in the present study.

The potential benefit of reducing the HAT down to the 4pHAT is twofold. Currently, the 10 puzzle HAT takes on average 18 minutes to complete (Burt, 2017). Reducing the measure to only four puzzles would likely bring the test time down to around 7 or 8 minutes. This shorter time should have the benefit of making the 4pHAT a more commercially viable product as organisations will be attracted to the shorter length of the test, meaning they will take less time assessing future candidates. Having the 4pHAT as an attractive tool is paramount to it addressing the current issues in workplace health and safety.

Criterion-Related Validity

In order to validate the HAT and 4pHAT as appropriate measures of hazard recognition, the present study needs to show that the two measures have criterion-related validity. Criterion-related validity refers to the extent to which one measure predicts scores on another measure (Salkind, 2010). In the present study, this means to what extent the HAT or 4pHAT predict the criterion measure of hazard recognition. The 'Hazard Lab' has been developed to measure actual hazard recognition of individuals. The aim is that if the 'Hazard Lab' has construct validity, that is, it measures what it claims to be measuring (hazard recognition) (Brown, 1996), then a relationship between the HAT or the 4pHAT and scores in the 'Hazard Lab' indicate criterion-related validity of the measures. Based on Schmidt and Hunter's (1998) meta-analysis on the validity of different selection methods, it would be expected that correlation between the predictor (HAT/4pHAT) and the criterion measure (Hazard Lab) should be around .37. Any result over .20 would lend to the moderate support of the HAT/4pHAT as a predictor of hazard recognition, while results over .50 would be very strong support (Schmidt & Hunter, 1998).

Current Study

Following the previous work done by Hill (2012), Shaw (2012), Thomas (2016) and Burt (2017) to create an objective tool that measures hazard recognition abilities, the current study aims to add to past research, with two main focuses. Firstly, a continued effort to further validate the HAT as a tool to measure hazard recognition abilities, and secondly, be the first study to validate the four-puzzle version of the HAT (4pHAT) as an independent tool. Given the self-report nature of previous validation studies on the HAT and the potential pitfalls that this may bring, the present study aims to validate both the HAT and the 4pHAT using a more objective criterion measure. To do this, a lab-based hazard recognition task was created so individuals actual hazard recognition abilities could be assessed in a controlled environment. As the tool is licenced to Talegent, it was used in conjunction with Talegent's own cognitive and personality measures. These two products have not been designed as measures of hazard recognition, but are measures to predict the general safety of candidates. Therefore, these tools, alongside the HAT/4pHAT acted as predictors of an overall safety criterion measure, created from in the form of an acquaintance questionnaire, as well as predictors of a combined criterion measure consisting of the overall safety measure and the hazard lab scores. The acquaintance questionnaire assessed the safety behaviour of participants through ratings on five safety components from a significant other of the participants. Below are the key hypotheses of the present study:

Hypothesis 1: Participants who perform better in the hazard lab hazard recognition task should have higher scores on the 4pHAT/HAT

a) Greater number of hazards identified

Hypothesis 2: Participants who score higher on the 4pHAT/HAT should show the following relationships with ratings collected from a significant other

a) High safety compliance
b) High safety participation
c) High safety consciousness
d) Low risk-taking ratings
e) High hazard recognition ratings

Hypothesis 3: The 4pHAT/HAT should account for unique variance over the cognitive and personality scores for the three criterion measures:

a) Hazard identification (Hazard Lab)
b) Overall safety (Acquaintance Questionnaire)
c) The combined measure (Hazard identification + Overall safety)

Method

Design Overview

The present study employed a predictive validation design to aid in the validation of the 4pHAt and the HAT.

Either the 4pHAT or the HAT was administered to participants alongside a cognitive abilities measure and a personality measure from Talegent, a firm specialising in psychometrics and measuring human performance. As well as providing further validation data for the present study, the commercial use of these tools as a package alongside the 4pHAT meant it was important to include them to keep continuity and to emulate real-world practices as much as possible. The administration of these tools was followed up approximately a week later by a *Hazard Recognition Task* (HRT) in the 'Hazard Lab' in which data was collected regarding *the number of hazards identified*, and the *time taken* to identify them. Data was also collected from an *acquaintance questionnaire* (AQ) (full AQ shown in Appendix A) completed by an acquaintance nominated by the participants completing the 4pHAT/HAT. The AQ asked questions related to the participant's *hazard recognition, safety contentiousness and risk-taking,* and *safety compliance and participation*.

To investigate the research question "Does the 4pHAT (and HAT) have criterion-related validity" 4pHAT or HAT scores were correlated with the five measures from the AQ and with the number of hazards identified and time take in the HRT.

The current investigation was reviewed and approved by the University of Canterbury Human Ethics Committee, reference number HEC 2018/62.

Participants

A sample of University of Canterbury undergraduate psychology students was used for the present study. Participants were recruited through the PSYC106 participant pool. This is a pool of students completing the Psychology 106: Introduction to Psychology course at the University of Canterbury who receive course credit for participation in one hour of approved departmental research. An advertisement (see Appendix B), giving a brief overview of the study, was posted on the participant pool online recruiting system where students would register for a participation time slot.

Acquaintances were recruited via the research participants. Participants were given an email with a link that would take their chosen acquaintance to the AQ. The link also contained an information sheet outlining the purpose of the research as well as a consent form that needed to be electronically completed for the AQ survey to open. Acquaintances were rewarded with a $10 petrol voucher for their participation in the research.

Table 2. shows the demographic breakdown of both participants and their acquaintances.

Table 2. Demographic information of participants by participation group

	4pHAT Participants	HAT Participants	Acquaintance Participants
Male	13	11	17
Female	22	24	36
Mean Age	21.34	23.77	32.23
(SD)	(6.53)	(11.85)	(15.31)
n	35	35	53

Materials - Hazard Awareness Test (HAT)

The HAT was presented as a computer-based task, run through the software package Eprime on a Desktop Windows 7 computer with a ViewSonic 1680 x 1050 display. Upon opening the program, participants were presented with instructions (Appendix C) on the screen, outlining how to complete the assessment. Participants are required to locate and click on, using a computer mouse, any differences they found between a picture pair (example shown in Figure 1. below). Each picture pair contained ten differences, five of which are safety related with the other five being neutral differences. All images were displayed in colour at a resolution of 95 dpi and to the dimensions of 1680 pixels wide and 1050 pixels in height. Along with the picture pairs, which were displayed in the centre of the screen, was an indication of the number of clicks remaining (out of 10) with a 'give up' button alongside. When a difference was correctly located, a green squared was displayed around the identified area, this acted as an indication for participants not needing to click in the given area for any more differences, as well as a way to track their progress. There was no time limit for the test. Images stayed onscreen until all the 10 differences were found or the participants clicked the "give up" button. Participants completed ten picture pairs in total. While there were no time restraints on taking the HAT, however, time take to complete the task was recorded.

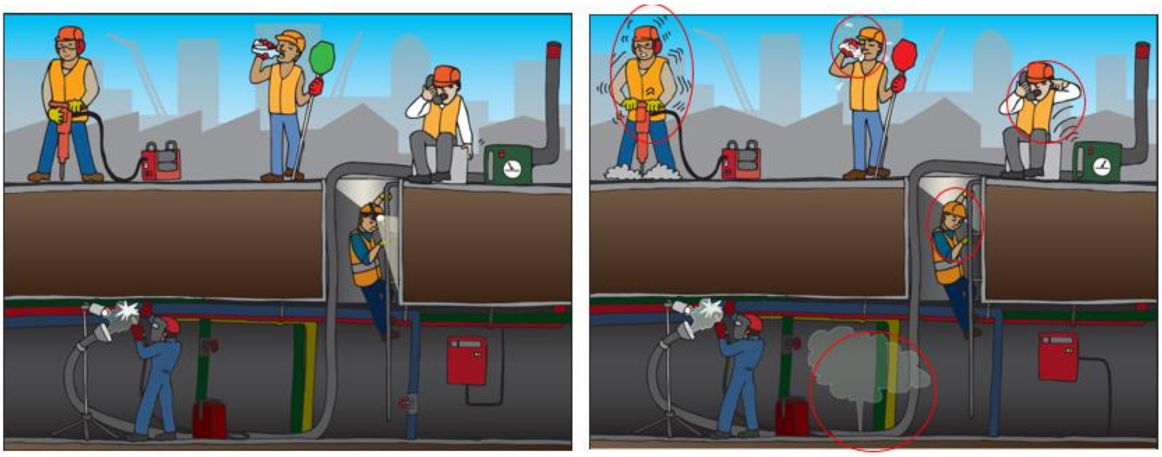


Figure 1. Example image pair from the HAT and 4pHAT

Hazard Awareness Test (HAT) Measures

Total number of safety differences found – This value is calculated as a sum of the safety-related differences found over the 10 puzzle pairs. As each pair has five safety-related differences the maximum score a participant could get is out of 50.

4pHAT

The *4pHAT*, for the purposes of this experiment, was run via Talegents web-based test administration site. Completion the of *4pHAT* is identical in every way to the HAT, other than participates only completing four picture pairs rather than ten that are used in the full HAT. As each pair has five safety-related differences the maximum score a participant could get is out of 20.

Talegent's Cognitive and Personality Measures

The cognitive measure for Talegent follows an information – question format. The questionnaire is headed with a short introduction, as shown in Appendix D. A short paragraph of information about a topic relating to health and safety is then presented to participants. This is followed by three multi-choice questions regarding the passage, each with four possible answers (an example question is shown in Appendix E). Participants have a maximum of one minute to answer each of the three questions. This process is repeated five times with different paragraphs of information each time.

The cognitive test was then followed by the personality measure. This measure included five individual variables; *Zero Harm Attitude, Dependable, Risk Conscious, Composed* and *Compliant.* Again, this was headed with a short introduction (Appendix F) prior to the questions being presented. Participants were then presented, eight at a time on each page, with statements in which they were to respond on a six-point scale from *Strongly Disagree* to *Strongly Agree*

These tools were developed to be utilised in a selection setting to give an indication of candidate's general safety behaviours such information retention from safety messages (cognitive measure) and their attitude to a 'zero harm' working environment (personality measure).

The Rating Criterion Variables: Acquaintance Questionnaire (AQ)

The AQ was run on Qualtrics, a web-based survey and questionnaire tool used through a licence with the University of Canterbury. The AQ was able to be completed on either a computer or a smartphone and was accessed through an anonymous link sent via an email to a chosen acquaintance by participants of the study.

Acquaintances were first presented with an information page outlining the purpose of the research and what was required of them to participate. They were also told at this point that they would be rewarded with a $10 Petrol Voucher for their participation. This was followed by a consent page where "I agree to participate in this research" needed to be selected in order to complete the questionnaire. If "I do not wish to participate in this research" was selected the survey terminated.

Following agreeing to participate, acquaintances were required to input a unique code given to them in the email from the participant in order for the data to be connected to the participant but still remain anonymous. Demographic information was then collected including; age, gender, relationship to the participants, length of said relationship and how well they knew the participant measured on a 100-point sliding scale from 0 (not very well at all) to 100 (extremely well).

Three randomised blocks of items are then presented to the acquaintances. The item blocks all have the instruction "*Listed below are a number of statements that could be used to describe *…'s safety behaviour. Please select the appropriate response by clicking on the corresponding circle to indicate how much you agree or disagree with each statement. If you don't know about any item please select the 'don't know' column.".* Acquaintances were instructed that the symbol (*…) pertains to the participant to which the questions are being answered about. The three scales (blocks of items) measure *hazard recognition*, *safety consciousness and risk-taking*, and *safety compliance and participation* are all responded to on a five-point scale from 1 '*Strongly Disagree'* to 5 '*Strongly Agree'* with a sixth option of '6' meaning *'Don't Know'.* For pre-existing scales being used modifications were made to the wording of the items to have them in third person format, rather than first person. For example "I put in extra effort to improve the safety of the workplace" became "*… puts in extra effort to improve the safety of the workplace....". Scores for the five individual scales (hazard recognition, safety compliance, safety participation, safety conscientiousness, risk-taking) were all created by averaging the recorded ratings across the respective scale items to create a single factor score. Scores of '6' (I don't know) were treated as 'missing data' so where not used in the calculation of the overall variable score, instead the average item score across all the participants was used to replace the 'missing data'. The individual variable scores could range from 1 – 5.

Hazard recognition was measured using an eight-item scale developed for this study. The development of the scale was conducted in conjunction with a review of the hazard recognition literature (Carter & Smith, 2006; Jeelani, Albert, Azevedo & Jaselskis, 2016; Kowalski-Trakofler & Barrett, 2003; Perlman, Sacks & Barak, 2014). The scale was designed to address the two main steps of hazard recognition, being sensation and knowledge. Items included *"*... is often aware when their physical surroundings are unsafe"* and *"*... understands when a situation is hazardous".* Cronbach's alpha of the scale for the present study was .83.

Safety compliance and participation were measured using sections of Neal and Griffin's (2006) scale designed to measure safety; climate, motivation, compliance and participation. Six items were used, with three for safety compliance and three for safety participation. Examples of the items include; for safety compliance *"*... always ensures the highest level of safety to carry out their job"* and *"*... promotes the safety program within the organization*" for safety participation. Ratings were averaged for the three respective items to give individual scores for both safety compliance and safety participation. Coefficient alphas, reported by Burt, Banks, and Williams (2014) were .93 and .86 respectively for safety compliance and safety participation. The present study reported coefficient alphas for safety compliance and participation of .83 and .76 respectively.

Safety consciousness and risk-taking were measured using Westby and Lee's (2003) 12-item scale developed to measure safety consciousness and risk-taking. Of the 12 items, seven measured safety consciousness while the remaining five measured risk-taking. Example items include *"*... takes a lot of time to do things safely even when it slows performance"* and *"*... loves to take risks even when there is a small chance *... could get hurt"* for safety consciousness and risk-taking respectively. Westaby and Lee (2003) reported a coefficient alpha of .77 and .85 for safety consciousness and risk-taking respectively in their initial study. In the present study coefficient alphas of .86 and .85 were reported for safety consciousness and risk-taking respectively.

The Behaviour Criterion Variable: Hazard Lab

The *Hazard Lab* is an 'office like' environment filled with 16 different 'hazards' for participants to identify and record. The 'office like' environment was chosen as it should be one familiar to the tertiary student sample. Having students through a 'construction site' environment would not have made for a fair test, as previous experience in this type of environment is not expected of tertiary students. Recording of the hazards formed the behavioural criterion variable for the validation study (along with the AQ data). While the hazards were made in a way to represent an actual risk in the environment, any real danger that the hazards appeared to pose was mitigated. This ensured that participants were not subjected to any real risk of physical harm. Table 4 below gives a description of each hazard present in the lab, along with how any real danger was mitigated and a picture of the hazard.

The room used contained a single entrance/exit point. From the perspective of the doorway, directly opposite the door in the far-left hand corner was a filing cabinet. Beside this, to the right was a desk with the broken computer chair in front of it, a computer on the desk and a whiteboard on the wall behind the desk. Along the right-hand wall of the room was a benchtop spanning the length of the room, as well as shelving above that, also running the entire length. Figure 2 shows a basic layout of the *hazard lab* with locations of all the hazards.

All hazards in the hazard lab were designed with the population sample in mind. It was necessary for all hazards to be recognisable by the student population. This meant that hazards were limited to what a student may come across in their day to day lives. While some of the hazards were quite difficult (i.e. expired fire extinguisher, and power cord tag) they could still be in the general environment of a student.

Two measures were taken from the *hazard lab; number of hazards identified,* and *time taken to ?.* Upon identifying a hazard, participants recorded a description of the hazard, potential harm it may cause, whether they believed it to be significant and how to rectify it, on a hazard register (Appendix G). Hazard registers were marked by the researcher against the descriptions in table xx. For example, if a participant recorded the heater as a hazard, but labelled it as a trip hazard rather than the fact the cover was missing, no mark was given. Scores for *number of hazards identified* could range from 0 – 16. Scores for *time taken* could have an unlimited range.

A small pilot study of the hazard lab was conducted prior to the beginning of the study. Pilot participants were also students from the University of Canterbury but were not currently enrolled in Psychology 106. This was done to keep a representative sample without disruption of the potential participant pool. Table 3. below shows the data obtained from the pilot. The *hazards identified* refers to the average number of the hazards, listed in Table 4, that were identified by the pilot sample. *Time taken* refers to the time average time the pilot sample spent in the ‘hazard lab’ identifying hazards.

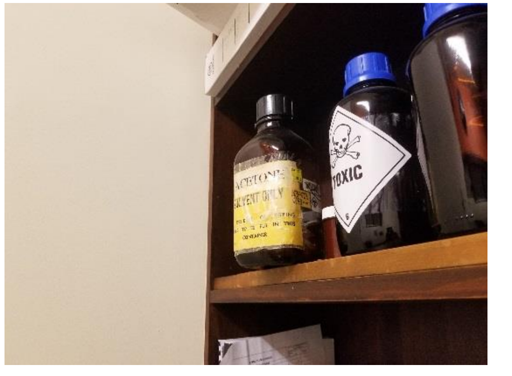
Table 3. *Summary data from Hazard Lab pilot study*

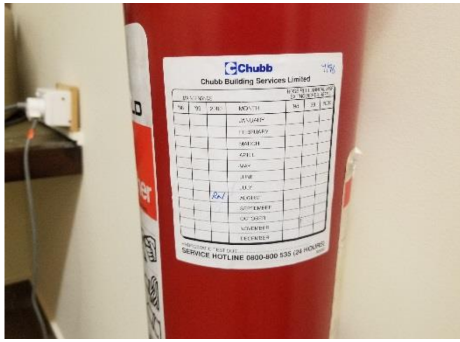
N	Hazards Identified (SD)	Time Taken (sec) (SD)
5	9.60 (2.07)	545.40 (224.81)

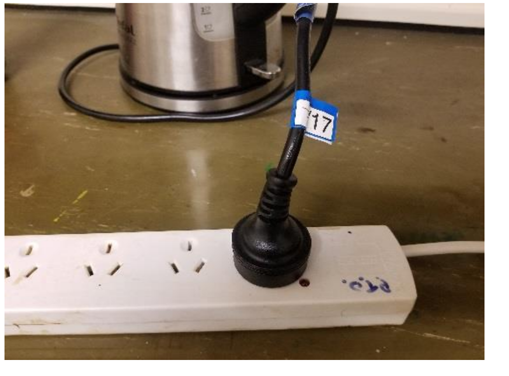
The *hazards identified* scores ranged from 6-13 suggesting that, even with the small sample, there does not appear to be any range restriction issues with the ‘hazard lab’ measure. Given an appropriate range in the data, it was concluded that the Hazard Lab was appropriate for use in the study.

Table 4. *List of hazards in the 'Hazard Lab', their descriptions, nullification of actual risk and photo*

	Hazards Name	Hazard Description	Nullification of Actual Risk	Photo of Hazard
1	*Coffee Spill*	Coffee spilt by a power multi-box where the jug is plugged in. Risk of electrocution.	No mains power being run, just a battery to power the light on the multi-box.	
2	*Glass*	Glass from a broken photo frame on the floor below as well as still in the frame. Potential to be cut by the glass.	The glass is sugar glass so poses no real risk with regards to cuts	
3	*Frayed Cable*	A lamp on the desk with a frayed and exposed power cable, risk of fire of electrocution.	No mains power being run, just a battery to power the lightbulb.	

4	*Power Socket*	A faulty power socket hanging from the wall with wires exposed, risk of electrocution.	No mains power being run through the socket. Is actually a mock socket	
5	*Chemical Jar*	A jar of hazardous chemicals perched on the edge of a shelf, both a potential falling hazard as well as a chemical hazard	The jar is attached to a pivot, holding it in place and is not able to fall. Also, does not contain any hazardous chemicals	
6	*Boxes*	Full boxes staked on high shelves, potential falling hazard.	Boxes are glued to the wall and are unable to fall on participants	

7	*Scissors*	Pair of scissors on the desk with a broken handle, the risk of cutting oneself if used.	Scissors glued to the desk, unable to be used	
8	*Smoke alarm*	Smoke alarm fixed to the wall with a flat battery, meaning it will fail upon a test	Room equipped with other fire protection	
9	*Fire Extinguisher*	Fire extinguisher fixed to the wall, expired and therefore useless in a fire	Room equipped with other fire protection	

10	*Extension Cord*	An extension cord laid across the floor with an anti-trip cover, not fulling covering the cable	The path to the cord blocked so the participant can’t actually trip	
11	*Broken Chair*	Chair at the desk with a broken leg, unable to be used without risk of falling off.	The chair cannot be pulled out from under the table.	
12	*Power Tag*	Cable for the Jug with an expired test tag, cable therefore potentially unsafe and an electrocution risk.	No mains power being run through the cable at all.	

13	*Microwave*	A metal bowl sitting inside of the microwave.	No mains power being run through the microwave, just a battery to power the clock.	
14	*Heater*	Small heater on the ground with a guard on the front missing.	No mains power being run through the heater. Made safe by tech support	
15	*Filling Cabinet*	Filling cabinet beside desk not appropriately fixed to the wall	The cabinet is actually fixed to the wall but can't be seen	

16	*Whiteboard Eraser*	Stuck on the whiteboard is a broken eraser with sharp plastic shards, potential to cut oneself.	No reason for the participant to use this – plus placed in the top corner of the whiteboard – out of reach.	

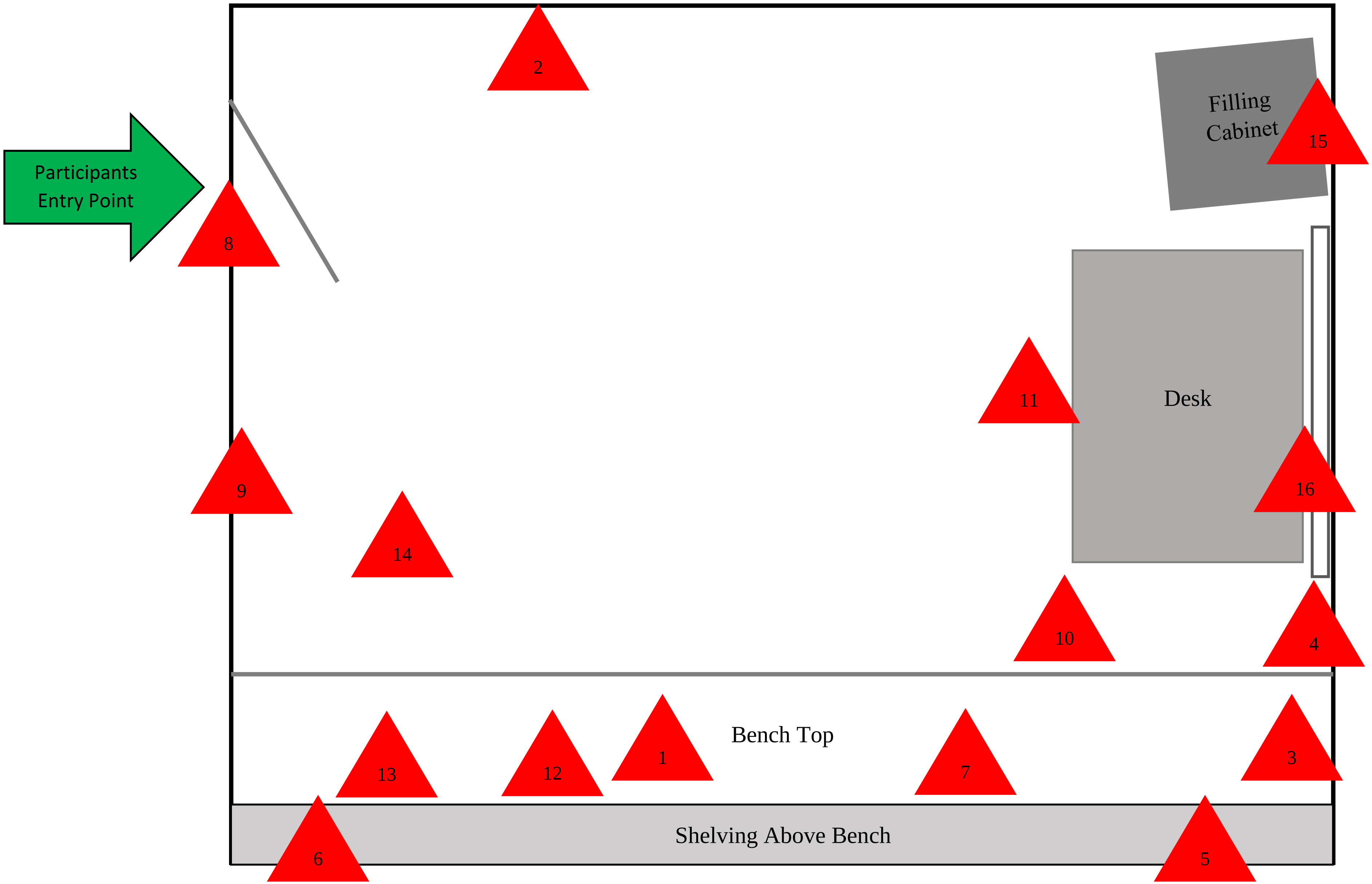


Figure 2. Layout of the 'Hazard Lab'

Procedure

The data, with the exception of the AQ, were collected under researcher supervision. Participants met the researcher outside Psychology Room 107 at the time agreed upon at sign up. Participants were first presented with an information sheet outlining the purpose of the research (Appendix H) and a consent form (Appendix I). Following an opportunity to ask questions and signing of the consent form participants were shown to a computer.

According to a random assignment, participants either completed the 4pHAT or the HAT, along with Talegents cognitive and personality measures. For those taking the 4pHAT, a link was open in a web browser showing Talegent's testing portal. Participant first filled out demographic information then preceded to take the 4pHAT, Cognitive Measure and Personality Measure, in that order. Participants taking the HAT did so on the E Prime software installed on the computer, then preceded to a link in the web browser containing only Talegent's cognitive and personality measures.

At the completion of these assessments, participants were emailed the link and instructions (Appendix J) for the AQ. Participants were told that the individual they chose to send the link to should be able to adequately answer questions on their safety behaviours, and were told that by sending the link, they were consenting to information being collected about them from a third party. A time was then agreed for them to return the following week to complete the second part of the research.

Upon returning to Psychology Room 107 approximately a week later at the agreed upon time, participants were given the following brief:

"*Shortly you will enter the office to complete an occupational health and safety hazard recognition register. Upon identification of a hazard, you are to; record a brief description, outline the potential harm, indicate the significance and how to deal with the* *hazard. Please use as many of the hazard register forms as you require. While there is no time limit to compete the register, your time taken will be recorded. When you have identified all the hazards, simply exit the room and hand your hazard register to the researcher".*

Participants were then given a clipboard with the Hazard Register form (Appendix G) and a pen. Following this they were shown into the 'Hazard Lab' and a timer was started. The door was closed behind them to mitigate any potential interference or influence from the researcher. While in the room participants would be identifying as many of the 16 hazards (as shown in Table 4. above) as they could.

Once participants exited the 'Hazard Lab' and handed the register to the researcher the timer was stopped and the time taken was recorded on the top of the hazard register.

Confirmation of the AQ being filled out for each participant was checked via Qualtrics and a $10 petrol voucher was given to the participant to pass on to their acquaintance if the data was there.

Finally, all participants were given as verbal and written debrief (Appendix K) outlining the purpose of the research and could ask any questions of the researcher.

Results

Data Preparation

The HAT, 4pHAT, hazard identification, cognitive test and *personality* data were added to an excel data file as it came in through the data collection process. Acquaintance questionnaire data was downloaded from the Qualtrics site at the conclusion of data collection. All data were transferred to SPSS for analysis.

Of the 70 participants involved in the study, six did not complete part two (i.e., the *Hazard Identification task,* and the return of an *Acquaintance Questionnaire).* This meant that only data pertaining to either the HAT or 4pHAT and the cognitive and personality assessments were available for analysis, for these 6 individuals. Of the remaining 64 participants, *acquaintance data* were not returned for 11 individuals. This meant that a total of 53 participants had acquaintance questionnaire data that could be used for analysis. Further to this, complete acquaintance questionnaire data were not available for a number of participants: due to excessive missing data one participant's *safety compliance* and *safety participation* scores, and two participant's *risk-taking* scores were not able to be calculated. Given the above information, different *n* = values will be evident throughout the results.

Furthermore, a small number of isolated missing values were also evident for scale items in the acquaintance questionnaire data. These missing item responses were substituted with the item mean value. Table 5 shows which items, the number of participants with a missing value, and the item mean used for the substitution process. Finally, the data were checked for outliers, defined as values plus or minus three standard deviations from the mean. No outliers were found in the data, and therefore all data, other than that mentioned above were used in the analysis.

Table 5. Missing acquaintance questionnaire items and the substituted means.

	Missing Item	Number Missing	Item Mean
Safety Consciousness	Item 1	1	3.94
	Item 2	1	3.87
	Item 3	1	3.78
	Item 6	1	3.92
Risk Taking	Item 9	3	3.17
	Item 11	4	3.37
Hazard Recognition	Item 3	1	4.41
Safety Compliance	Item 1	1	4.26
Safety Participation	Item 4	4	4.09
	Item 5	3	3.98
	Item 6	2	3.92

Variable Distribution and Range Restriction Issues

Validation analysis can be adversely affected by range restriction in either the predictor or the criterion variables. Range restriction has the effect of suppressing relationships in the correlation-based analysis (Raju & Brand, 2003). Analysis was conducted examining the descriptive statistics, and the distribution of the data to identify any range restriction issues. Skewness and Kurtosis were analysed to determine the ability of the data to be used for correlational analysis, which ideally uses normally distributed data (Hunter, Schmidt, & Le, 2006). For evidence of a normally distributed data set with a sample size of 50 or less, absolute Z values for both skewness and kurtosis should be less than 1.96 (Kim, 2013). These Z values are calculated by dividing the actual skewness value by the standard error. For samples sizes between 50 and 300, values over 3.29 constitute a non-normal sample distribution (Kim, 2013).

Predictor Variables

Results pertaining to the mean, standard deviation, range, skewness and kurtosis of the predictor variables are shown in Table 6.

Table 6. Mean, range, skewness and kurtosis statistics of the predictor variables.

	n	Mean (SD)	Range	Skewness (SE)	Skewness Z-score	Kurtosis (SE)	Kurtosis Z-score
HAT	35	46.14 (3.68)	30-50	-2.66 (.40)	-6.65	10.24 (.79)	12.96
4pHAT	35	18.17 (1.99)	11-20	-1.79 (.40)	-4.48	3.89 (.79)	4.92
Cognitive test	70	28.87 (3.58)	17-34	-1.44 (.29)	-4.97	2.14 (.57)	3.75
Zero Harm Attitude	70	88.54 (8.44)	66-107	-.27 (.29)	-.93	-.01 (.57)	-.02
Dependable	70	44.96 (6.23)	30-60	.26 (.29)	.90	-.25 (.57)	-.44
Risk Conscious	70	42.49 (11.39)	19-72	.62 (.29)	2.14	.21 (.57)	.37
Composed	70	41.96 (7.76)	27-58	.08 (.29)	0.28	-.59 (.57)	-1.04
Compliant	70	50.99 (18.06)	28-100	1.51 (.29)	5.21	1.24 (.57)	2.18

Table 6. indicates that the *HAT, 4pHAT, Cognitive* assessment, and one dimension of the personality data (compliant) have skewness and kurtosis values that indicate non-normally distributed data (Kim, 2013). The respective large negative skewness values of -2.66 (*SE=.40), -*1.79 *(SE=.40), -*1.44 *(SE= .29) and 1.51 (SE.29)* indicate an asymmetry of the data with a long tail to the left of the distribution with the bulk of the data lying to the

right-hand side of the mean. Kurtosis expresses the degree to which the density of the data differ from what is expected with a normal distribution (Hopkins & Weeks, 1990). The respective, large kurtosis values of 10.24 *(SE= .79)*, 3.89 *(SE= .79)* and 2.14 (*SE= .57)* indicate leptokurtic distributions with a large peak to the data.

There are two major factors as to why the predictor data present in this way. The *HAT*, *4pHAT* and *Cognitive* measures are all ability tests. That is, they measure an individual's ability on a specific dimension. In the case of the *HAT* and *4pHAT*, the measurement is of hazard recognition ability. Data skewed to the upper end of the distribution is somewhat expected for this type of measure. The assumption is that, in general, individuals have some level of hazard recognition ability, likely learnt through experience, otherwise a much larger number of people would be having accidents more frequently.

Figures 3 and 4 show a plot of the scores on the *HAT* and *4pHAT*, respectively. The x-axis represents the individual cases, while the y-axis represents their score on the respective measures. The imposed red line identifies the mean *HAT* and *4pHAT* scores respectively.

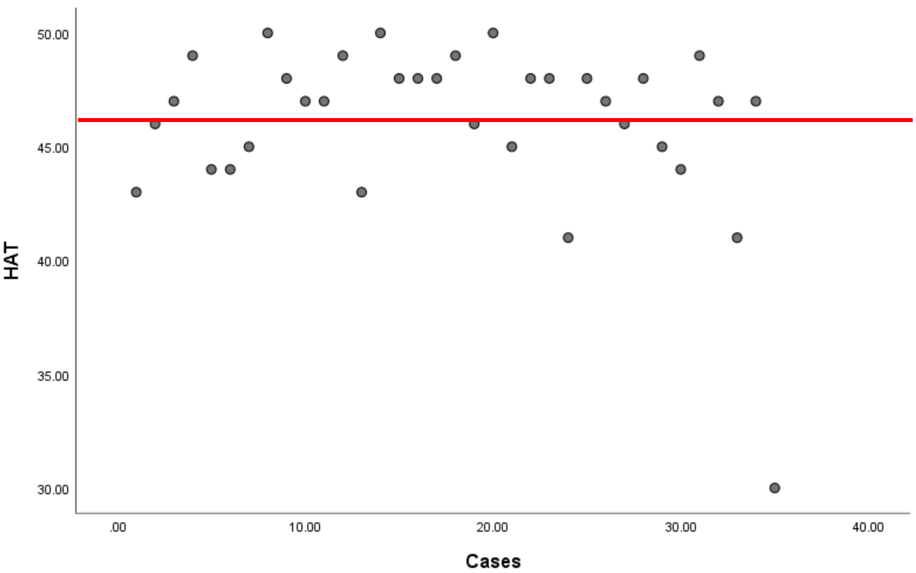


Figure 4. Plot of all cases scores on the HAT with mean score imposed in red.

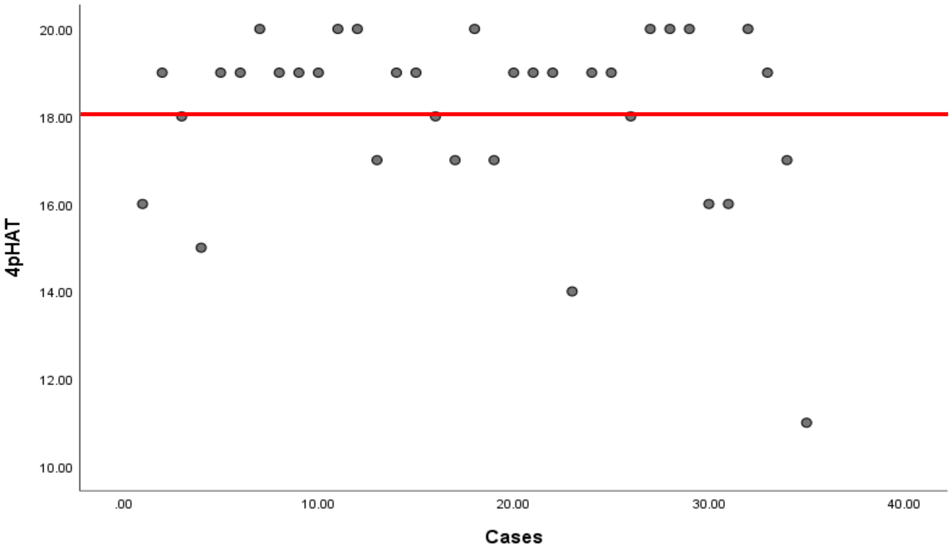


Figure 5. Plot of all cases scores on the 4pHAT with mean score imposed in red.

As the main purpose of the *HAT* and *4pHAT* is to identify those individuals that have a poor level of hazard recognition ability, it is necessary for the measures to pick up individual differences. Inspection of Figures 3 and 4 indicate that the measures do in fact identify individual differences, with six scores on the *4pHAT* and three scores on the *HAT* sitting outside one standard deviation. This shows that while the bulk of the data grouped together, the measures are still able to identify individuals with lower hazard recognition abilities.

Regarding the *cognitive* measure, given the bulk of participants (*n=61*) were university students, it can be assumed that they may have higher cognitive ability than the general population. This would account for the skewness 1.44 *(SE=.29)* and kurtosis 2.14 *(SE=.57)* of the data for *cognitive* ability measure.

The above-discussed range restriction issues should not have any practical influence on the correlation analysis. Range-restricted data causes suppression of correlations (Cascio & Aguinis, 2008). This suggests that correlations in the following analysis may be less than would be expected with a normally distributed data set.

Criterion Variables

The objective criterion variables for the present study are the two hazard lab measures; *hazards identified,* and *time taken.* Further criterion measures come from the acquaintance questionnaire. These data consist of five safety measures: *safety compliance, safety participation, safety consciousness, risk-taking* and *hazard recognition*. Assessment of these measures, with reference to range issues, is displayed in Table 7. The items for *risk-taking* have been reverse coded so that all variables are in a positive direction and are more favourable regarding safety behaviour (i.e. high *risk-taking* scores now mean fewer risks would be taken by an individual). To make this apparent in the analysis *Risk Taking* will be referred to as *Less Risk Taking.*

Table 7. Mean, range, skewness and kurtosis statistics of the criterion variables.

	n	Mean (SD)	Range	Skewness (SE)	Kurtosis (SE)
Hazard Lab					
Hazards Identified	64	10.17	3-14	-.559	-.142
		(2.62)		(.30)	(.59)
Time Taken (Seconds)	64	953.78	390-1964	.767	-.008
		(378.52)		(.30)	(.59)
Acquaintance Data					
Hazard Recognition	53	4.37	3.13-5	-.436	-.679
		(.48)		(.33)	(.64)
Safety Conscientiousness	53	3.94	2.29-5	-.332	-.605
		(.68)		(.33)	(.64)
Less Risk Taking	51	3.40	1.40-5	-.255	-.774
		(.90)		(.33)	(.66)
Safety Participation	52	4.01	2.26-5	-.026	-.675
		(.62)		(.33)	(.65)
Safety Compliance	52	4.21	2.33-5	-.511	-.344
		(.68)		(.33)	(.65)

Inspection of the results in Table 7 indicate no range restriction issues with the criterion variables. It is therefore assumed that all measures are normally distributed.

Criterion-related validity: Objective Hazard lab data

The main goal of the present study was to assess the criterion-related validity of the HAT and the 4pHAT. Pearson product moment correlations were calculated to assess the relationship between the *HAT* and the *4pHAT* and the two hazard lab measures (i.e., *hazards identified, time taken).* The obtained correlations, as well as the means and standard deviations, are shown in Table 8.

Table 8. Descriptive statistics, and Pearson correlation of the 4pHAT and HAT with two hazard lab measures.

	4pHAT (n=30)	Mean (SD)	HAT (n=34)	Mean (SD)
Hazard Lab (Hazards Identified)	.53**	10.30 (2.82)	.46**	10.06 (2.47)
Hazard Lab (Time in seconds)	.29	926.97 (361.02)	.07	977.44 (397.19)

Note: ** $p < .01$

The results shown in Table 8 indicates large positive correlations for both the *HAT* and the *4pHAT* with the number of hazards identified in the hazard lab (*hazards identified).* These results support the criterion-related validity of the two measures and indicate that individuals who score higher on the *HAT* or *4pHAT* also identified a greater number of hazards in the hazard lab. Review of Table 8 also suggests there is no significant relationship between *HAT* or *4pHAT* scores and the *time taken* in the hazard lab.

Criterion related validity: Acquaintance data

Pearson correlation analysis was also calculated between the *HAT, 4pHAT* and the five measures from the acquaintance questionnaire; *safety compliance, safety participation, safety consciousness, risk-taking* and *hazard recognition,* as well as an *overall safety* measure. An overall safety measure was also calculated by summing the scores for the five acquaintance questionnaire measures. The *overall safety* measure has a mean of 19.93 (SD=2.85). The correlations, and means and SDs are shown in Table 9.

Table 9. Person Correlations of acquaintance data scores with the HAT and 4pHAT.

	n	4pHAT	n	HAT
Safety compliance	23	.23	29	-.25
Safety participation	23	.34	29	-.15
Safety consciousness	23	$.43^{*}$	30	-.11
Less Risk taking	22	.18	29	.05
Hazard recognition	23	$.43^{*}$	30	-.21
Overall Safety	22	.37	28	-.12

Note: * $p < .05$

Inspection of Table 9. shows significant positive correlations between the *4pHAT* and the *safety consciousness* and *hazard recognition* measures. These results indicate that individuals who scored higher on the *4pHAT* were rated higher on *hazard recognition* and *safety consciousness* behaviours by their acquaintance. No further correlations were found to be significant. However, the correlations between *safety compliance, safety participation* and the *overall safety measure* and the *4pHAT* are positive and large in terms of criterion-related validity analysis. However, the small sample means the values are not significant. Table 9. also shows contrasting correlations between the HAT (where mainly negative correlations are evident) and the 4pHAT were mainly positive correlations were found. These results are from different groups of participants who were randomly assigned to a test condition, and potential explanations for the differing results will be addressed in the discussion section of this paper.

Incremental Validity Analysis

In a real-world operational sitting the *HAT* or the *4pHAT* will be used alongside other measures when job applicants are being assessed. A key issue is whether each of the measures used adds incrementally to the prediction of the applicant's safety behaviour. For the present study, the HAT and 4pHAT are considered in relation to Talegent's safety-related *personality* and *cognitive* measures.

One potential issue with the current data set, regarding running a regression analysis is the small sample size. A small sample may cause the regression results to be unstable and give inappropriate and inaccurate values (Kelley & Maxwell, 2003). In an attempt to address this the *4pHAT* (n= 35) and *HAT* (n= 35) were combined to give a larger sample (n= 70) to work with. This was done by standardising both scores to a percentage (i.e *4pHAT/*20 and *HAT/*50) and thus creating one variable which represented hazard identification assessment.

To address the question of incremental validity regression analysis was run against several criterion measures, including the *hazard identification* scores, the combined *overall safety* measure from the acquaintance questionnaire data, and a *combined criterion* measure. The combined criterion variable represented all the safety data obtained on participants- that which represented their measured behaviour, and which represented others views of their safety behaviour. Utilising this measure is the most holistic way to capture an individual's safety behaviours as it included both their direct ability as well as what others perceive of their safety behaviours. The combined criterion measure was created by summing the *hazard identification* and *overall safety* measures. The *combined criterion* measure has a mean of 30.27 (*SD*=4.10), a skewness value of -.08 (*SE=.34*) and kurtosis of -.79 (*SE=.66*) indicating no range restriction issues (Kim, 2013).

Prior to the regression analysis, the predictor variables were investigated for any potential multicollinearity. This was done via a correlational analysis being run between both the *HAT* and *4pHAT* individually (as well as with the *standardised Hazard recognition ability* score), *Cognitive* and *Personality,* as shown in Table 10.

Table 10. Person Correlations and descriptive statistics of cognitive and personality scores with the 4pHAT, HAT and Standardised HAT

	4pHAT (n=30)	HAT (n=34)	Standardised HAT (n=70)
Cognitive	.28	.19	.25*
Zero Harm Attitude	.32	-.10	.14
Dependable	.23	-.27	.00
Risk Conscious	.01	-.03	-.02
Composed	.15	.12	.14
Compliant	.19	.07	.14

Note: * $p < .05$

Inspection of Table 10 shows one significant correlation between the *cognitive* measures and the *Standardised HAT* (p=.04). This small positive correlation suggests a relationship between the assessment scores, but not one significant enough to have an effect on the regression analysis. However, while these results suggest no multicollinearity between the predictor variables further collinearity analysis will be conducted during the regression analysis, by examining the variance inflation factor (VIF).

Three separate regressions were run. The first was against the objective *hazard identification* measure. The second was against the acquaintance reported *overall safety* measure. The final regression was against the combined criterion measure. The regressions predicting *hazard identification* [F(7, 56) = 4.82, p = .00], *overall safety* [F(7, 42) = 2.78, p = .02], and the *combined criterion* [F(7, 42) = 6.76, p = .00] all produced significant models. Further details of the regression results are shown in Table 11.

Table 11. Combined regression analyses of the *cognitive*, *personality* and *standardised HAT* measures against the *hazard lab*, *overall safety* and the *combined criterion* measure

	Hazard Lab (n=63)			Overall Safety (n=49)			Combined Criterion (n=49)		
	β	SE	*p*	β	SE	*p*	β	SE	*p*
(Constant)	-9.27*	4.30	.04	4.95	5.98	.41	-10.36	7.14	.15
Cognitive	.00	.08	.99	.01	.11	.93	-.03	.13	.83
Zero Harm Attitude	.03	.04	.80	.21	.05	.13	.21	.06	.08
Dependable	-.00	.05	.99	.14	.07	.39	.13	.08	.33
Risk Conscious	.27*	.03	.02	.39**	.03	.01	.44**	.04	.00
Composed	.26*	.04	.04	-.13	.05	.39	.05	.06	.70
Compliant	-.03	.02	.80	.03	.02	.81	-.01	.03	.93
Standardised HAT	.46**	.03	.00	.10	.05	.49	.45**	.06	.00
Adjusted R^2	.30**			.20*			.45**		

Note: ** $p < .01$, * $p < .05$

Collinearity analysis was conducted with all three regressions. VIF values did not exceed 1.65, thus it is concluded multicollinearity is of no issue for the present data (Schroeder, Lander & Levine-Silverman, 1990). Inspection of Table 11 identifies several key findings. Firstly, the full assessment package including the *standardised HAT, personality* and *cognitive* measures significantly predict all three of the criterion measures, with the package accounting for 45% of the variance in the *combined criterion* measure. Secondly, the s*tandardised HAT* adds a unique and significant contribution to predicting both the *hazard lab* and the *combined criterion* but does not significantly add to predicting the *overall safety* measure. The final key finding is *risk-conscious* adding a unique and significant contribution to predicting all three criterion measures, as well as *composed* predicting *hazard lab* scores.

Discussion

Study Aims

The aim of the present study was twofold; to investigate the criterion-related validity of the Hazard Awareness Test (HAT), and to independently validate the more commercially viable version of the tool, the 4pHAT. Along with the main research aim, the validity of 4pHAT being used in conjunction with Talegent's cognitive and personality measures was assessed. The continued validation of these tools aimed to address the gap in the market for a valid and objective health and safety measure to be utilized in the selection process of applicants to high-risk workforces.

Summary of Finding – Hypotheses

The criterion-related validity of the two measures was assessed through their relationship with the 'hazard lab' task and ratings from an acquaintance about the participant's health and safety behaviours. Hypothesis 1 investigated whether those who performed better on the 'hazard lab' hazard recognition task would also have larger scores on the 4pHAT/HAT. The results showed statistically significant positive correlations between both the 4pHAT and the HAT with the *number of hazards identified* in the 'hazard lab' (see Table 8), lending support to Hypothesis 1a, which in turn provides support for the criterion-related validity of both the 4pHAT and the HAT. The significance of this results is on the size of the criterion-related validity represented by the correlation. A conclusion around reasonable criterion-related validity would have been drawn based on Schmidt and Hunter's (1998) review, with a value of .37 being the target of the present study. However, the present study has resulted in values in the region of .50. These values put the 4pHAT and the HAT in line with the validity scores for General Mental Ability (GMA) tests and Work Sample tests, seen by Schmidt and Hunter's review as the most effective predictive tools.

Hypothesis 2 investigate whether larger scores on the 4pHAT/HAT would be associated with larger ratings across the five factors from the acquaintance questionnaire. Results showed mixed support for Hypothesis 2. The relationships with the 4pHAT across all five of the acquaintance questionnaire (AQ) factors, as well as the *overall safety* composite score, showed small to medium positive correlations (see Table 9). These results suggest that scores on the 4pHAT could be related to greater reported safety behaviours. While the 4pHAT itself is not a measure of general safety behaviour, the ability to provide a unique contribution to the measurement of a holistic safety measure (such as the *combined criterion* shown in Table 11) in conjunction with Talegent's other measures, increases its practical applications. The results also included a significant positive correlation between the 4pHAT and the *hazard recognition* factor in the AQ, further supporting the criterion-related validity of the 4pHAT as a tool to measure hazard recognition ability. The results pertaining to the HAT, however, showed some puzzling findings. Small negative correlations were found between the HAT and all the AQ factors, including the *overall safety* composite, other than no relationship being found for *less risk taking.* These conflicting results will be discussed further in a later section.

Hypotheses 3 investigated the unique variance the 4pHAT/HAT had over Talegent's cognitive and personality measures when predicting the three criterion measures. Utilising the combined *standardised HAT* score, to account for the low sample size, the results suggested strong support for Hypothesis 3a with the *standardised HAT* score providing a significant and unique contribution to predicting *hazard lab* scores (see Table 11). No support was found for Hypothesis 3b with the *standardised HAT* score not providing a unique contribution to the prediction of *overall safety.* Support for Hypothesis 3c was shown with the standardised HAT score providing a significant and unique contribution to predicting the overall *combined criterion* measure. When considered alongside the criterion-related validity evidence in the support of hypothesis 1 and 2, the support in hypotheses 3 suggests the usefulness of the 4pHAT/HAT to be used in a selection setting to measure hazard recognition ability, in conjunction with Talegents measures. The unique contribution made by the 4pHAT/HAT in predicting not only hazard recognition, but also a combined measure of safety, means that the three measures together should account for a greater measurement of an individuals safety behaviours.

Summary of Finding – Other findings

The general trend thought the data showed sometimes opposite results, as mentioned above, between the relationships shown across many variables and the HAT/4pHAT. As discussed, the relationship between both the HAT and the 4pHAT with the *number of hazards identified* in the 'hazard lab' is consistent across the two predictors (see Table 8). However, across many the other variables, both at the predictor end (see Table 10) and the criterion end (see Table 9), correlations are very conflicting, in some cases opposite. There are a number of reasons this may have occurred. With a relatively small sample size of 35 for each tool, individual difference will have a greater impact on the results than if the sample was larger. This could have created the inconsistent results shown as it was different people in each condition. This could have been most apparent in the acquaintance questionnaire data. Given the strong and very similar correlations between the 4pHAT/HAT and the *hazards identified*, it seems less likely that the variance in the participants had a big effect. However, the data gathered from the participant's acquaintances would also be prone to variance from individuals difference and may have had more of a significant effect on the results. This is highlighted when comparing the results in Tables 9 and 10. Table 9 contains data collected from the acquaintance questionnaire and shows vastly differing results across all six of the variables. Comparing this to the results in Table 10, containing the data from the Talegent measures, only two of the five factors (Zero Harm Attitude, Dependable) have results that a vastly different between the HAT and the 4pHAT. From this, it is concluded that the most likely explanation of the differing results comes from individual difference variation from the participant's acquaintances in the acquaintance questionnaire data.

While no specific hypotheses were developed for Talegent's cognitive and personality measures, there is some value in a discussion of the results shown for these measures. Firstly, it should be noted that neither of the measures was specifically designed to be measures of hazard recognition. They were designed for use in a selection setting to measure more general safety behaviours of candidates. Of Talegent's predictors, the *risk-conscious* factor provided the most significant contribution to prediction all three criterion measures. For lack of a given definition, risk-conscious is an individual's state of awareness to potential dangers in their environment (APA, n.d.). In the context of hazard recognition, it seems clear that a relationship would exist between an individual's state of awareness for dangers and the first step in the hazard recognition process of sensing the hazard (Jeelani, Albert, Azevedo & Jaselskis, 2016). The *risk-conscious* factor was a significant predictor of the *number of hazards identified* in the 'Hazard lab' (see Table 11) suggesting support to the relationship outlined by the discussed definition above. While the Talegent's measures will be used alongside the 4pHAT regardless, due to the licencing agreement, there does seem to be some valid benefit in utilizing this measure to account for further variance in prediction hazard recognition ability.

It should be noted that the evidence presented above also suggest the validity of the 4pHAT/HAT used independently of Talegent's measures.

Practical and Theoretical Implications

The main premise behind the present study lies with reducing the high prevalence of workplace accidents through appropriate safety measures. If at the recruitment and selection stage, organisations can identify individuals who may be lacking in the safety behaviours and abilities required to safely perform their jobs, we could see a significant reduction in workplace accidents. The proven ability of the HAT and the 4pHAT to effectively predict individuals hazard recognition abilities, though an objective, and valid, gamified tool, provides organisations with a way to measure one such of these safety abilities, in hazard recognition. The most likely real-world application of these measures is in high-risk industries, where a high level of hazard recognition ability is required. Therefore, the main aim of the tool should not be to identify individuals who have an exceptionally high level of hazard recognition but should be to identify those individuals who do not have a well-developed hazard recognition ability (as shown in figures 4 and 5). These tools should be utilized in a selection space to identify those candidates that may not have an appropriate level of hazard recognition and therefore may pose a health and safety risk to an organisation. Given hazard recognition is the first step in the safety management process (Perlman, Sacks & Barak, 2014), the impact of selecting individuals with an appropriate level of hazard recognition could have a significant effect on reducing the rate of incidents.

While having these effective tools available is a step in the right direction, it is necessary for the use of these tools to be attractive to organisations, as the tools have no impact sitting on the shelf. The present studies validation of the 4pHAT could conceivably increase this attractiveness. While the time taken to complete the 4pHAT was not recorded in the present study, at only 40% the length of the HAT, it is assumed that completion time on the 4pHAT could be 60% less. Based on Burt's (2017) completion time of the HAT being on average 18 minutes, the 4pHAT would take approximately seven minutes on average to complete. Given time is everything in a commercial environment, if a tool can be administered and completed in less time, while still providing the same level of measurement accuracy and effectiveness, it would be of great benefit to an organisation. The 4pHAT compares well to the Wonderlic test, one of the most extensively used measures of problem-solving ability and IQ (Murphy, 1984), regarding the time to complete the test. At only 12 minutes to complete, the major attraction of the Wonderlic is in its short administration time (Wonderlic & Associates, 1983) compared to other measures of similar abilities such as the Raven's test which takes more than 40 minutes to complete (Raven, Court & Raven, 1985). It is hoped that the 4pHAT will benefit in a similar way from its short completion time.

The present study adds to the theories discusses in two main areas, gamification and safety/accident theory. As disused earlier, the literature is short of work conducted in the workplace selection and assessment space. While the present study does not specifically focus on the effectiveness of gamification principles in selection tools, it does provide an example of them in an applied setting. No definitive conclusions can be drawn between the gamified nature of the tool and its proven effectiveness as a measurement tool, however, given the HAT was specifically designed with gamification principles in mind, the findings show some support to the idea of utilising gamified measurement tools in the selection space.

Regarding safety and accident theory, the present study does not provide a direct contribution to the literature but rather provides an avenue for future research to support the current theories. Hazard recognition is proposed in the literature as a key stage in safety management (Perlman, Sacks & Barak, 2014) and accident causation (Carter & Smith, 2006). Using the 4pHAT or the HAT, researchers have an objective and valid measure to measure the hazard recognition abilities of individuals, which they can intern relate to an accident that has occurred, to confirm the importance of this ability.

Limitations

The results of the present study should be viewed with consideration of the potential limitations of the analysis.

The sample size of the present study posed a limitation to the appropriateness of the analysis conducted. For the correlational analysis, the sample size of 35 for each of the 4pHAT and the HAT was appropriate, however, this sample was not large enough to conduct regression analysis with. To deal with this the samples were compiled into one measure to create a larger sample of 70. While this sample size would lend itself to a more stable regression, it is potentially still too small to draw any concrete conclusions from. Based on Green's (1991) analysis of the sample size needed for multiple regressions, the present study would require a sample size of over 100 to show substantial results. For the purpose of the study, the most important results came from the criterion-related validity evidence, assessed through the correlational analysis. The regression analysis was used to assess the unique contribution the tools had over the Talegent measures.

Both the sample used, and the testing environment provide a potential challenge in concluding generalisability of the results to the greater target population. The sample consisting of mostly young, university students (average age = 22.56 years) likely only represents a small percentage of the age and education of the target population. Those in target industries, such as construction and agriculture, likely have not attended university, and have instead taken up a trade.

The testing environment likely provides the greatest limitation to the present analysis. The rigorously controlled, artificial 'hazard lab' environment may have an effect on an individual hazard recognition. As participants are taking part in a university ethics approved study, they would have been safe in the knowledge that they were under no real danger. Any apparent 'danger' represented by the hazards in the 'hazard lab' was mitigated for the safety of the participants. In a real-world environment where there is a real risk of danger and therefore a chance of increased stress, individual ability to identify hazards may be severely affected. The stressful environment would likely increase their ability to identify those hazards which are more life threating, as they would naturally want to be preserving their life but may reduce their ability to identify those hazards which may not present an immediate danger.

Future Research

Future research has been discussed to some degree with the theoretical implication of the present study, however, there is another key area that would benefit from more study. To address the limitations of the sample in the present study, there may be a benefit in further validation of the 4pHAT, utilizing a job-specific sample, and in a real-world setting. This could be done in a more controlled experimental setting with a ‘hazard lab’ type environment being set up on an actual construction site and running present employees through the experiment. Conversely, validation could be conducted concurrently through workers who had been employed utilizing the 4pHAT as a selection tool. Actual accident reports that included a hazard recognition component could be measured against the scores obtained on the 4pHAT at the selection phase.

Conclusion

The present study aimed to validate the use of the 4pHAT and the HAT as tools to measure hazard recognition ability. Through the utilisation of predictive validation method in a controlled, lab-based, hazard recognition task context, the evidence presented supports the validly of the two tools. While further research would be beneficial to validate the tools in an appropriate work sample, the present study has taken a step to fill the void in the health and safety measurement space with a valid, and objective, gamified tool. It is hoped that the shorter 4pHAT tool will be seen by safety-conscious organisations as an attractive tool that can have a significant impact in reducing the number of workplace accidents, and the subsequent human and financial cost.

References

Albert, A., Hallowell, M. R., & Kleiner, B. M. (2013). Enhancing construction hazard recognition and communication with energy-based cognitive mnemonics and safety meeting maturity model: Multiple baseline study. *Journal of Construction Engineering and Management*, 140(2).

Albert, A., Hallowell, M. R., Skaggs, M., & Kleiner, B. (2017). Empirical measurement and improvement of hazard recognition skill. *Safety science*, 93, 1-8.

APA Dictionary of Psychology. (n.d.). Retrieved from https://dictionary.apa.org/risk

APA Dictionary of Psychology. (n.d.). Retrieved from https://dictionary.apa.org/conscious

Arnold, H. J., & Feldman, D. C. (1981). Social desirability response bias in self-report choice situations. *Academy of Management Journal*, 24(2), 377-385.

Attali, Y., & Arieli-Attali, M. (2015). Gamification in assessment: Do points affect test performance?. *Computers & Education*, *83*, 57-63.

Bahn, S. (2013). Workplace hazard identification and management: The case of an underground mining operation. *Safety Science*, 57, 129-137.

Barrett, P. (2010). A review of commercial products and academic articles associated with the psychological assessment of safety attributes within prospective and incumbent employees. Retrieved from: http://www.pbarrett.net/Safety_Review/A_Review_of_the_Psychological_Assessment_of_Safety_Attributes_Paul%20Barrett_Oct_2010.pdf

Brown, J. D. (1996). *Testing in language programs.* Upper Saddle River, NJ: Prentice Hall Regents.

Buckley, P., & Doyle, E. (2016). Gamification and student motivation. *Interactive Learning Environments*, 24(6), 1162-1175.

Burt, C. D. B., Banks, M. & Williams, S, (2014). The safety risks associated with helping others. *Safety Science*, 62, 136-144.

Burt, C. D. B., (2015). *New Employee Safety: Risk Factors and Management Strategies*. Springer.

Burt, C. D. B., (2017). *Hazard Awareness Test Description and Validation Evidence Manual*. Unpublished Manuscript. University of Canterbury, Christchurch, New Zealand.

Carter, G., & Smith, S. D. (2006). Safety hazard identification on construction projects. *Journal of Construction Engineering and Management*, 132(2), 197-205.

Cascio, W.F. & Aguinis, H. (2011). *Applied psychology in human resource management* (7th Edition). Prentice Hall.

Chamorro-Premuzic, T., Winsborough, D., Sherman, R. A., & Hogan, R. (2016). New talent signals: Shiny new objects or a brave new world?. *Industrial and Organizational Psychology*, 9(3), 621-640.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4), 349.

Csikszentmihalyi, M. (1991) *Flow: The psychology of optimal experience*. Harper Perennial.

DeJoy, D. M., Schaffer, B. S., Wilson, M. G., Vandenberg, R. J., & Butts, M. M. (2004). Creating safer workplaces: assessing the determinants and role of safety climate. *Journal of Safety Research,* 35(1), 81-90.

de Klerk, S., Veldkamp, B. P., & Eggen, T. J. (2015). Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a Bayesian network example. *Computers & Education*, 85, 23-34.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). *From game design elements to gamefulness: defining gamification.* In Proceedings of the 15th International academic Mind Trek conference: Envisioning future media environments. (pp. 9-15). ACM.

DuVernet, A. M., & Popp, E. (2014). Gamification of workplace practices. *The Industrial-Organizational Psychologist, 52*(1), 39-44.

Eickhoff, C., Harris, C. G., de Vries, A. P., & Srinivasan, P. (2012, August). *Quality through flow and immersion: gamifying crowdsourced relevance assessments*. In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval (pp. 871-880). ACM.

Feyer, A., & Williamson, A. (2003). *Occupational injury: Risk, prevention, and intervention*. London; Bristol, PA; Taylor & Francis.

Flin, R., Mearns, K., O'Connor, P., & Bryden, R. (2000). Measuring safety climate: Identifying the common features. Safety science, 34(1-3), 177-192.

Fiore, C A. (2016). Gamification of Individual Differences. *Culminating Projects in Psychology*. Paper 4.

Gangadharbatla, H., & Davis, D. Z. (Eds.). (2016). Emerging research and trends in gamification. *Information Science Reference.*

Green, S. B. (1991). How many subjects does it take to do a regression analysis. *Multivariate Behavioural Research*, 26(3), 499-510.

Hamari, J., Koivisto, J., & Sarsa, H. (2014, January). *Does gamification work?--a literature review of empirical studies on gamification*. In 2014 47th Hawaii international conference on system sciences (HICSS) (pp. 3025-3034). IEEE.

Health and Safety at Work Act. (2015). Retrieved from http://www.legislation.govt.nz/act/public/2015/0070/latest/whole.html#DLM6375600

Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist.* 78-79.

Hill, A. (2012) *Workplace Safety Orientation: The Construct Validation of a New Measure* (Masters dissertation, University of Canterbury, Christchurch, New Zealand). Retrieved from http://ir.canterbury.ac.nz/handle/10092/7716

Hopkins, K. D., & Weeks, D. L. (1990). Tests for normality and measures of skewness and kurtosis: Their place in research reporting. *Educational and Psychological Measurement*, 50(4), 717-729.

Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, 91(3), 594.

Huotari, K., & Hamari, J. (2012, October). *Defining gamification: a service marketing perspective*. In Proceeding of the 16th international academic MindTrek conference (pp. 17-22). ACM.

Independent Taskforce on Workplace Health and Safety (2013). *The Report of the Independent Taskforce on Workplace Health & Safety*. Retrieved from http://hstaskforce.govt.nz/documents/report-of-the-independent-taskforce-onworkplace-health-safety.pdf

Jeelani, I., Albert, A., Azevedo, R., & Jaselskis, E. J. (2016). Development and testing of a personalized hazard-recognition training intervention. *Journal of Construction Engineering and Management*, 143(5), 04016120.

Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. John Wiley & Sons.

Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, 8(3), 305.

Kim, H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, 38(1), 5

King, M. F., & Bruner, G. C. (2000). Social desirability bias: A neglected aspect of validity testing. *Psychology & Marketing*, 17(2), 79-103.

Koepp, M. J., Gunn, R. N., Lawrence, A. D., Cunningham, V. J., Dagher, A., Jones, T., . . . Grasby, P. M. (1998). Evidence for striatal dopamine release during a video game. *Nature,* 393(6682), 266–268

Kowalski-Trakofler, K. M., & Barrett, E. A. (2003). The concept of degraded images applied to hazard recognition training in mining for reduction of lost-time injuries. *Journal of Safety Research*, 34(5), 515-525.

Kumar, J. (2013, July). *Gamification at work: Designing engaging business software*. In International conference of design, user experience, and usability (pp. 528-537). Springer, Berlin, Heidelberg.

Lamm, F. (2002). Occupational health and safety in small businesses. Occupational health and safety in New Zealand: Contemporary social research.

Laughery, K. R., & Vaubel, K. P. (1989). The role of accident experiences on subsequent accident events. *Occupational injury: Risk, Prevention and Intervention*, 33-43.

Levashina, J., & Campion, M. A. (2007). Measuring faking in the employment interview: development and validation of an interview faking behavior scale. *Journal of applied psychology*, *92*(6), 1638.

Levashina, J., Hartwell, C.J., Morgenson, F.P., & Campion, M.A. (2014). The structured employment interview: Narrative and Quantitative review of the research literature. *Personnel Psychology*, 67, 1, 241-293.

Levy, R. (2013). Psychometric and evidentiary advances, opportunities, and challenges for simulation-based assessment. *Educational Assessment*, 18(3), 182-207.

Lowman, G. H. (2016). Moving beyond identification: Using gamification to attract and retain talent. *Industrial and Organizational Psychology*, 9(3), 677-682.

Maudgalya, T., Genaidy, A. & Shell, R. (2008). Productivity–quality–costs–safety: A sustained approach to competitive advantage—a systematic review of the national safety council's case studies in safety and productivity. *Human Factors and Ergonomics in Manufacturing & Service Industries*. 18(2), 152-179.

Menezes, C. C. N., & De Bortolli, R. (2016). Potential of Gamification as Assessment Tool. *Creative Education*, 7(04), 561.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (Vol. 821). John Wiley & Sons.

Murphy, K. R. (1984). The Wonderlic personnel test. *Test critiques*, 1, 769-775.

Namian, M., Albert, A., Zuluaga, C. M., & Jaselskis, E. J. (2016). Improving hazard recognition performance and safety training outcomes: Integrating strategies for training transfer. *Journal of Construction Engineering and Management*, *142*(10),

Neal, A., Griffin, M. A., & Hart, P. M. (2000). The impact of organizational climate on safety climate and individual behavior. *Safety Science*, 34(1-3), 99-109

Neal, A., & Griffin, M. A. (2006). A study of the lagged relationships among safety climate, safety motivation, safety behavior, and accidents at the individual and group levels. *Journal of Applied Psychology*, 91(4), 946.

Perlman, A., Sacks, R., & Barak, R. (2014). Hazard recognition and risk perception in construction. *Safety Science*, 64, 22-31.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, 12(4), 531-544.

Probst, T. M., Barbaranelli, C., & Petitta, L. (2013). The relationship between job insecurity and accident under-reporting: A test in two countries. *Work & Stress*, 27(4), 383-402.

Raju, N. S., & Brand, P. A. (2003). Determining the significance of correlations corrected for unreliability and range restriction. *Applied Psychological Measurement*, *27*(1), 52-71.

Ramsey, J. D. (1985). Ergonomic factors in task analysis for consumer product safety. *Journal of Occupational Accidents*, 7(2), 113-123.

Ramsey, J. D. (1989). Assessment of warnings based on an ergonomic accident sequence model. *International Journal of Industrial Ergonomics*, 4(3), 195-199.

Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. *Safety Science*, 27(2-3), 183-213.

Raven, J. C., Court, J. H., & Raven, J. (1985). *A manual for Raven's Progressive Matrices and Vocabulary Scales*. London: H. K. Lewis.

Salkind, N. J. (2010). *Encyclopaedia of research design.* Thousand Oaks, CA: SAGE Publications, Inc.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*(2), 262.

Schroeder, M. A., Lander, J., & Levine-Silverman, S. (1990). Diagnosing and dealing with multicollinearity. *Western Journal of Nursing Research*, 12(2), 175-187.

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, 74, 14-31.

Shappell, S. A., & Wiegmann, D. A. (2000). The human factors analysis and classification system--HFACS.

Shaw, E. (2012) *Use of spot the difference puzzles as a measure of occupational safety orientation* (Masters dissertation, University of Canterbury, Christchurch, New Zealand). Retrieved from http://ir.canterbury.ac.nz/handle/10092/7648

Shrout, P. E., & Yager, T. J. (1989). Reliability and validity of screening scales: Effect of reducing scale length. *Journal of Clinical Epidemiology*, 42(1), 69-78.

Sijtsma, K. (2012). Future of psychometrics: Ask what psychometrics can do for psychology. *Psychometrika*, 77(1), 4-20.

Westaby, J. D., & Lee, B. C. (2003). Antecedents of injury among youth in agricultural settings: A longitudinal examination of safety consciousness, dangerous risk taking, and safety knowledge. *Journal of Safety Research*, 34(3), 227-240.

Wilson, H. A. (1989). "Organizational behaviour and safety management in the construction Industry." *Construction Management and Economics*, 7(4), 303–319.

Wonderlic, E. F., & Associates (1983). *Wonderlic Personnel Classification Test manual*. Northfield, IL.

Wood, R., & Payne, T. (1998). *Competency-Based Recruitment and Selection*. Wiley.

WorkSafe New Zealand. (2017a). Health and Safety at Work Act 2015. Retrieved from https://worksafe.govt.nz/laws-and-regulations/acts/hswa/

WorkSafe New Zealand. (2017b). Introduction to the Health and Safety at Work Act 2015 – special guide. Retrieved From: https://worksafe.govt.nz/managing-health-and-safety/getting-started/introduction-hswa-special-guide/

Young, J. (2012). *Operational risk management: The practical application of a qualitative approach*. Pretoria, South Africa: Van Schaik.

Appendix A: *Acquaintance Questionnaire*

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Validation of the 4pHAT: Criteria Validity Evidence

Acquaintance Information Sheet

My name is Jason Hunt and I am a Masters of Applied Psychology student at the University of Canterbury conducting a study to measure the validity of the 4pHAT (A subset of the Hazard Awareness Test). The purpose of the research is to establish if the 4pHAT is a valid measure of individuals hazard recognition abilities.

If you choose to take part in this study, your involvement in this project will be to spend approximately 10 minutes completing the following acquaintance questionnaire. This questionnaire includes several safety and hazard identification behaviour items about *.................. who has invited you to complete this questionnaire. Whenever you see *... below this refers to the person who invited you to participate in this study. For your time you will be remunerated with a $10 Petrol voucher, given to *..... once you have competed this survey and we have your responses recorded.

Participation is voluntary and you have the right to withdraw at any stage without penalty. You may ask for your raw data to be returned to you or destroyed at any point. If you withdraw, I will remove information relating to you. However, once analysis of raw data starts on [date to be filled in after ethics is accepted], it will become increasingly difficult to remove the influence of your data on the results.

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation: your identity will not be made public without your prior consent. To ensure confidentiality no identifying information is collected in the survey and responses are anonymous. No one other than me (as the researcher) and Chris Burt (as the research supervisor) will have access to the data. Electronic data will be stored on a password-protected computer, in a locked room. Data will be destroyed after 5 years unless a publication outlet requires extended archiving of the data. A thesis is a public document and the subsequent thesis will be available through the UC Library.

A separate link at the end of the survey will allow you to enter your email address if you would like a copy of the research findings

The project is being carried out as a requirement for the Masters of Applied Psychology programme by Jason Hunt under the supervision of Chris Burt, who can be contacted at christopher.burt@canterbury.ac.nz. He will be pleased to discuss any concerns you may have about participation in the project.

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee, and participants should address any complaints to The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).

If you agree to participate in the study, you are asked to agree to the consent information following, before completing the acquaintance questionnaire.

→

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Validation of the 4pHAT: Criteria Validity Evidence
Acquaintance Consent Form

I have been given a full explanation of this project.

I understand what is required of me if I agree to take part in the research.

I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.

I understand that any information or opinions I provide will be kept confidential to the researcher and that any published or reported results will not identify the participants. I understand that a thesis is a public document and will be available through the UC Library.

I understand that all data collected for the study will be kept in locked and secure facilities and/or in password-protected electronic form and will be destroyed after five years, unless a publication outlet requires extended archiving of the data.

I understand there are no risks associated with taking part in this study

I understand that I can contact the researcher [Jason Hunt jason.hunt@pg.canterbury.ac.nz] or supervisor [Chris Burt christopher.burt@canterbury.ac.nz] for further information. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz)

By indicating below, I agree to participate in this research project.

- I agree to participating in this research
- I do not wish to participate in this research

→

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Please input the code supplied by the individual who invited you to participate

Gender

- Male
- Female
- Gender Diverse
- Prefer not to say

Age

How do you know *...

- Work Colleague
- Work Manager
- Work Supervisor
- Spouse
- Child
- Parent
- Friend
- Sport/Recreation associate
- Other

How long have you known *...

- Years
- Months

Please indicate how well you know *…. by placing a mark on the 100 point scale.

Not Very Well At All									Extremely Well	
0	10	20	30	40	50	60	70	80	90	100

Safety Consciousness

Listed below are a number of statements that could be used to describe *…'s safety behaviour. Please select the appropriate response by clicking on the corresponding circle to indicate how much you agree or disagree with each statement. If you don't know about any item please select the 'don't know' column.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Don't Know
*… always take extra time to do things safely	○	○	○	○	○	○
People think of *… as being an extremely safety-minded person	○	○	○	○	○	○
*… always avoid dangerous situations	○	○	○	○	○	○
*… takes a lot of time to do things safely even when it slows performance	○	○	○	○	○	○
*… often makes sure that other people do things that are safe and healthy	○	○	○	○	○	○
*… gets upset when seeing other people acting dangerously	○	○	○	○	○	○
*… thinks doing the safest possible thing is always the best thing	○	○	○	○	○	○
*… would rather take risks than be overly cautious	○	○	○	○	○	○
In the past month *… has done some exciting things that other people might think are dangerous	○	○	○	○	○	○
*… loves to take risks even when there is a small chance *… could get hurt	○	○	○	○	○	○
Sometimes people get on *… nerves when they tell *… how to act "more safely"	○	○	○	○	○	○
*… values having fun more than being safe	○	○	○	○	○	○

Hazard Recognition

Listed below are a number of statements that could be used to describe *…'s safety behaviour. Please select the appropriate response by clicking on the corresponding circle to indicate how much you agree or disagree with each statement. If you don't know about any item please select the 'don't know' column.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Don't Know
*... is often aware when their physical surroundings are unsafe	○	○	○	○	○	○
*... points out hazardous situations	○	○	○	○	○	○
*... is always aware of their physical surroundings	○	○	○	○	○	○
*... is an observant individual	○	○	○	○	○	○
*...knows a hazardous situation when they see one	○	○	○	○	○	○
*... has good situational awareness	○	○	○	○	○	○
* ... notices physical changes in familiar environments	○	○	○	○	○	○
*... understands when a situation is hazardous	○	○	○	○	○	○

Safety Behaviour

Listed below are a number of statements that could be used to describe *…'s safety behaviour. Please select the appropriate response by clicking on the corresponding circle to indicate how much you agree or disagree with each statement. If you don't know about any item please select the 'don't know' column.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Don't Know
*... always uses all the necessary safety equipment	○	○	○	○	○	○
*.... promotes the safety programme within their organisation	○	○	○	○	○	○
*... always uses the correct safety procedures	○	○	○	○	○	○
*... always ensure the highest level of safety	○	○	○	○	○	○
*... voluntarily carries out tasks or activities that help to improve workplace health and safety	○	○	○	○	○	○
*... puts in extra effort to improve the safety of their workplace	○	○	○	○	○	○

→

UC UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Would you like a copy of the research findings emailed to you at the conclusion of the research?

- (•) Yes
- ( ) No

ENTER EMAIL ADDRESS HERE

→

UC UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

English ▾

Please enter your email address below if you would like a copy of the research findings

>>

Appendix B: *Recruitment Advertisement*

Study Name	Gamification Fun - Validation of the 4pHAT
Study Type	**Multi-Part Study** This is a Multi-Part study. There are 2 parts. All parts must be signed up for at the same time. Part 2 should be scheduled to occur between 5 and 10 day(s) after Part 1 Part 2 may be scheduled to occur at any time on a different day than Part 1 and that is within the range of acceptable dates.
Study Status	**Visible to participants :** Approved **Active study :** Appears on list of available studies
Duration	30 minutes (Part 1) 20 minutes (Part 2)
Credits	1 Credits (Part 1) 1 Credits (Part 2) (2 Credits total)
Abstract	The present study is being conducted to validate the 4pHAT, a Health and Safety Selection Tool that identifies individuals hazard recognition abilities.
Description	Participants will be required to complete two 30minutes session. The first will consist of three short computer based tasks and the second will involve a lab based task. Participants will also need a significant other (family, colleague, friend) to fill out a short questionnaire that will be emailed to them. They will receive a $10 Petrol Voucher for their time.

Appendix C: *HAT Instructions*

Instructions

You will see 10 pairs of images displayed side-by-side. Your task is to find the differences between the two images in each pair.

There are a total of 10 DIFFERENCES in each image pair.

You have a total of 10 ATTEMPTS (mouse clicks) to find the differences. Please move the computer mouse cursor over the difference on the RIGHT-HAND image, and click on the difference with the left-hand mouse button. If a difference is correctly identified a green indicator box will be displayed. If an error occurs, the selected area will not be highlighted.

It is important to place the middle of the mouse cursor (+) directly over the difference before clicking.

Once your 10 attempts to find the differences in an image pair are completed you will be moved on to the next test page. If you are stuck and cannot find more differences, please click the Give Up button to move onto the next test page.

Appendix D: *Talegents' Cognitive Instructions*

Safety Instruction Comprehension

On the screens that follow you will be presented with passages of text containing safety information. Each passage will have **three** questions about the information and you will be allowed **1 minute** to answer as many of the three questions as possible.

IMPORTANT: You should answer the questions based on the information provided, rather than drawing on your own experience.

This section is timed. You may not be able to answer all questions in this section so please don't worry if you can't finish in the time provided.

To help prepare you for this section, you will be presented with an example passage of information with three un-scored questions. When you are ready to begin please click the **next** button.

Appendix E: *Talegents' Cognitive Example*

Example Question

Do not overload fuses on alternators. Always adhere to personal protective equipment standards of clothing and safety equipment – goggles, shields, lenses, gloves, and leather jacket – at all times. Always check that all equipment is well maintained, except plugs and cords, which are checked yearly by a qualified electrician.

How often are the plugs and cords checked?

- A) Once a week
- B) Once a year
- C) Once a month
- D) Daily

Appendix F: *Talegents' Personality Instructions*

How do you view rules and procedures?

This section assesses your inclinations toward guidelines. On the screens that follow are statements that may or might not describe you. Please read each statement carefully, thinking about it in a work context. Decide whether you tend to agree or disagree with it. Choose one answer that best corresponds to your level of agreement or disagreement.

There are no right or wrong answers; your answers help us understand how you prefer to act at work. Although this section is **untimed**, you should try to work reasonably quickly. Please click the next button when you are ready to begin.

Please click the NEXT button when you are ready to begin.

Appendix G: *Hazard Register*

HAZARD REGISTER

Code ____________

#	Description of Hazard	Potential Harm	Significant Hazard?		Eliminate	Isolate	Minimise
			Yes	No			

Appendix H: *Information Sheet for Participants*

Department of Applied Psychology
Email: Jason.hunt@pg.canterbury.ac.nz

25.05.2018

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Validation of the 4pHAT
Information Sheet for Participants

My name is Jason Hunt and I am a Masters of Applied Psychology student at the University of Canterbury conducting a study of the validity of the 4pHAT (A subset of the Hazard Awareness Test (HAT)). The purpose of the research is to establish if the 4pHAT is a valid measure of individuals hazard recognition abilities.

If you choose to take part in this study, your involvement in this project will be required to complete three short computer based tasks consisting of a cognitive measure, personality measure and either the HAT of the 4pHAT. All together the tasks should not take more than 30 minutes to complete. At the completion of these tasks you will be emailed a link to a survey to be completed by an acquaintance of your choosing, reporting on your hazard recognition and safety behaviors. **You are to send this link, along with a unique code given to you by the researcher, to your chosen acquaintance and in doing so you are consenting to them giving data about your behaviours.**

As a follow-up to this investigation, you will be asked to return to the lab and complete a This process will take no more than 30 minutes. You will also receive a $10 Petrol Voucher to give to your acquaintance, provided they have filled out the survey by this time. This will be known by referencing the survey responses with code given to you in time one.

Participation is voluntary and you have the right to withdraw at any stage without penalty. You may ask for your raw data to be returned to you or destroyed at any point. If you withdraw, I will remove information relating to you. However, once analysis of raw data starts on [*insert date at which withdrawal of data, to be complete after ethics*], it will become increasingly difficult to remove the influence of your data on the results.

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation: your identity will not be made public without your prior consent. To ensure anonymity and confidentiality, your name will only be present on a consent form kept separate from all collected data which will be coded for anonymity. All raw data will only be view by myself, as the researcher and Chris Burt, as the research supervisor. Physical data will be stored in a locked filing cabinet in a locked room. Electronic data will be stored on a password protected computer, in a locked room. Data will be destroyed after 5 years, unless a publication outlet requires extended archiving of the data. A thesis is a public document and will be available through the UC Library.

Please indicate to the researcher on the consent form if you would like to receive a copy of the summary of results of the project.

The project is being carried out as a requirement for the Masters of Applied Psychology programme by Jason hunt under the supervision of Chris Burt, who can be contacted at christopher.burt@canterbury.ac.nz. He will be pleased to discuss any concerns you may have about participation in the project.

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee, and participants should address any complaints to The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).
If you agree to participate in the study, you are asked to complete the consent form and return to the researcher prior to commencing participation.

Appendix I: *Consent Form for Participants*

Department of Applied Psychology
Email:
Jason.hunt@pg.canterbury.ac.nz

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Validation of the 4pHAT
Consent Form for Participants

- ☐ I have been given a full explanation of this project and have had the opportunity to ask questions.
- ☐ I understand what is required of me if I agree to take part in the research.
- ☐ I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.
- ☐ I understand that any information or opinions I provide will be kept confidential to the researcher and that any published or reported results will not identify the. I understand that a thesis is a public document and will be available through the UC Library.
- ☐ I understand that all data collected for the study will be kept in locked and secure facilities and/or in password protected electronic form and will be destroyed after five years, unless a publication outlet requires extended archiving of the data.
- ☐ I understand there are no risks associated with taking part in this study.
- ☐ I understand that by sending the survey link to my chosen acquaintance I consent to information about me being gathered from said acquaintance
- ☐ I understand that I can contact the researcher [*Jason Hunt* jason.hunt@pg.canterbury.ac.nz] or supervisor [*Chris Burt* christopher.burt@canterbury.ac.nz] for further information. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz)
- ☐ I would like a summary of the results of the project.
- ☐ By signing below, I agree to participate in this research project.

Name:___________________________Signed:___________________________Date:____________

Email address *(for report of findings, if applicable)*:

____

Please return this consent form to the researcher prior to commencing participation.

Appendix J: *Acquaintance email the link and instructions*

Dear Acquaintance,

You have been selected (by the individual who forwarded you this email) to participate in a research study being conducted at the University of Canterbury.

Participation requires you to answer a number of questions pertaining to the individuals Health and Safety Behaviours. You are not at all obliged to participate, though your participation would be greatly appreciated.

Click the link below to participate – where you will find an Information sheet, consent form and the questionnaire itself.

http://canterbury.qualtrics.com/jfe/form/SV_8Jt72g3YS8z6IPr

For your time and participation, you will receive a $10 petrol voucher.

Please ensure you input the following unique code on the questionnaire so your responses can be linked to the participant who nominated and you can receive your voucher. The voucher will be given to the individual who nominated you for collection

Unique Code 10XX

If you have any questions, feel free to contact the researcher at jason.hunt@pg.canterbury.ac.nz

Kind Regards,

Jason Hunt

Appendix K: *Participant Debrief*

Participant Debrief -Validation of the 4pHAT

The present study aims to validate the 4pHAT, a subset of the Hazard Awareness Test (HAT). The HAT and 4pHAT are Health and Safety Selection tools, used to measure individuals Hazard Recognition ability. It is hoped that the 4pHAT and the HAT can be used in organisations to reduce the number of workplace accidents resulting from a lack of hazard recognition by screening for this ability at the recruitment and selection phase.

Previous validation has been carried out of the HAT. However, the present validation will be more specific to the 4pHAT and will be utilising criterion measures that better reflect actual hazard recognition abilities.

Multiple criterion measures are being used to aid in this validation such as the number of hazards participants are able to identify, the time taken to find those hazards, and ratings of hazard recognition and safety behaviours given by an acquaintance of the participants.

It is hypothesised that the criterion measures listed above will be positively correlated with the scores on the 4pHAT.

It is asked that you do not share your knowledge or experience of this experiment with other potential participants until the completion of data collection on the 1st of November.

Please feel free to ask and questions or express any concerns with the researcher.

Thank you for participating in this research.

Jason Hunt
jason.hunt@pg.canterbury.ac.nz